

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/688,019;
; FILING DATE: 19-APR-1991
; APPLICATION NUMBER: 08/004,199
; FILING DATE: 13-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConti, Giulio A., Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-006CNCPP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 953 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..681
US-08-197-793-1
```

```
Query Match 0.7%; Score 22; DB 1; Length 953;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3079 tgtgataaaataaaataaa 3100
DB 903 TGTGATAAAATAAAATAAA 924
```

```
RESULT 11
US-08-636-176-1
; Sequence 1, Application US/08636176
; Patent No. 584822
; GENERAL INFORMATION:
; APPLICANT: Meuer, S.
; APPLICANT: Schraven, B.
; APPLICANT: Schoenhaut, D.
; APPLICANT: Ratnofsky, S.
; TITLE OF INVENTION: pp32: A Newly Identified CD45-Associated
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/636,176
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/688,019; 08/004,199
; FILING DATE: 19-APR-1991; 13-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConti, Giulio A., Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-006CNCPP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 953 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..681
US-08-636-176-1
```

```
Query Match 0.7%; Score 22; DB 2; Length 953;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3079 tgtgataaaataaaataaa 3100
DB 903 TGTGATAAAATAAAATAAA 924
```

```
RESULT 12
PCT-US95-01618-1
; Sequence 1, Application PC/TUS9501618
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: pp32: A Newly Identified CD45-Associated
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01618
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/197,793
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConti, Giulio A., Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-006CPCP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 953 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..681
PCT-US95-01618-1
```

```
Query Match 0.7%; Score 22; DB 5; Length 953;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3079 tgtgataaaataaaataaa 3100
|||||
```

Db 903 TGTGAAAAAAAAAAAAAAAA 924

RESULT 13

US-08-884-324-2

; Sequence 2, Application US/08884324

; Patent No. 6060283

; GENERAL INFORMATION:

; APPLICANT: Takanori OKURA

; APPLICANT: Kakuji TORIGOE

; APPLICANT: Masahi KURIMOTO

; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE

; TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/884,324

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 185,305/96

; FILING DATE: 27-JUN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BROWDY, Roger L.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: OKURA-1

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1120 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna to mRNA

; HYPOTHETICAL: No

; ANTI-SENSE: No

; ORIGINAL SOURCE:

; ORGANISM: human

; TISSUE TYPE: liver

; FEATURE:

; NAME/KEY: 5'UTR

; LOCATION: 1..177

; IDENTIFICATION METHOD: E

; NAME/KEY: leader peptide

; LOCATION: 178..285

; IDENTIFICATION METHOD: S

; NAME/KEY: mat peptide

; LOCATION: 286..756

; IDENTIFICATION METHOD: S

; NAME/KEY: 3'UTR

; LOCATION: 757..1120

; IDENTIFICATION METHOD: E

US-08-884-324-2

Query Match 0.7%; Score 22; DB 3; Length 1120;

Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3079 tgtgaaaaaaaaaaaaaaaa 3100

Db 1099 TGTGAAAAAAAAAAAAAAAA 1120

RESULT 14

US-08-832-180-6

; Sequence 6, Application US/08832180

; Patent No. 6214584

; GENERAL INFORMATION:

; APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU

; APPLICANT: KENYUJO

; APPLICANT: USHIO, Shimeji

; APPLICANT: TORIGOE, Kakuji

; APPLICANT: TANIMOTO, Tadao

; APPLICANT: OKAMURA, Haruki

; APPLICANT: KURIMOTO, Masashi

; TITLE OF INVENTION: INTERFERON-

; TITLE OF INVENTION: POLYPEPTIDE

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Browdy and Neimark

; STREET: 419 Seventh Street N.W. Ste. 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM: disk

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect Version 5.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/832,180

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/558,191

; FILING DATE:

; APPLICATION NUMBER: JP 304,203/94

; FILING DATE: No. 6214584ember 15, 1994

; APPLICATION NUMBER: 10048102

; FILING DATE: September 18, 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Browdy, Roger L.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: USHIO-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 628-5197

; TELEFAX: (202) 737-3528

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1120 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna to mRNA

; HYPOTHETICAL: No

; ANTI-SENSE: No

; ORIGINAL SOURCE:

; ORGANISM: human

; TISSUE TYPE: liver

; FEATURE:

; NAME/KEY: 1-177 5'-UTR

; IDENTIFICATION METHOD: S

; NAME/KEY: 178-285 leader peptide

; IDENTIFICATION METHOD: S

; FEATURE:

; NAME/KEY: 286-756 mat peptide

; IDENTIFICATION METHOD: S

; FEATURE:

; NAME/KEY: 757-1120 3'-UTR

; IDENTIFICATION METHOD: S



---

OM of: US-09-647-377-9 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Jul 8, 2002 2:12 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09647377/runat\_08072002\_080021\_23088/app\_query.fasta\_1.940  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=olip2n.rni  
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LCOPEL=0.000  
-LOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000  
-XGAPEXT=60.000 -GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=60.000  
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -STARTP=1  
-MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200  
-THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09647377@cgn1\_1.142 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-647-377-9  
Query length: 870  
Database: Issued\_Patents\_NA:\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 55.110000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:	Strd Orig	zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-743-637B-192 +	8.00	117.39	74.91	732	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-247-155-44 +	8.00	115.96	90.01	893	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-853-1 -	8.00	114.94	102.60	1029	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-911-853-1 -	8.00	114.94	102.60	1029	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-409-1 -	8.00	114.94	102.60	1029	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-453-1 -	8.00	113.32	126.33	1289	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-358-627F-22 +	8.00	113.08	130.22	1332	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-814A-3 -	8.00	112.35	143.07	1475	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-282-305-3 +	8.00	112.28	144.33	1489	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-836-582-1 -	8.00	112.28	144.33	1489	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-265-566-1 -	8.00	110.81	174.25	1826	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-282-305-1 +	8.00	108.31	240.21	2585	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-579-777A-1 +	8.00	107.01	283.84	3097	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-228-986-11 +	8.00	104.52	390.67	4377	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-853-28 -	8.00	104.52	390.67	4377	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-409-28 -	8.00	104.52	390.67	4377	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-453-28 -	8.00	89.53	2.7e+03	35060	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-814-095-7 -	7.00	117.50	73.89	59	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-270-805C-18 -	7.00	117.50	73.89	59	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-410-654B-18 -	7.00	117.50	73.89	59	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-474-851-18 -	7.00	117.50	73.89	59	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-560-18 -	7.00	117.50	73.89	59	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-643-810A-18 -	7.00	117.03	78.50	63	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-270-805C-17 +	7.00	117.03	78.50	63	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-410-654B-17 +	7.00	117.03	78.50	63	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-474-851-17 +	7.00	117.03	78.50	63	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-560-17 +	7.00	117.03	78.50	63	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-170-113-10 +	7.00	117.03	78.50	63	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-643-810A-17 +	7.00	117.03	78.50	63	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-552-613-10 +	7.00	114.96	102.39	84	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-731-157A-34 +	7.00	114.96	102.39	84	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-541-780-34 +	7.00	113.77	119.16	99	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-752B-10 +	7.00	112.39	142.33	120	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-403-676A-96 -	7.00	112.39	142.33	120	
/cgn2_6/ptodata/2/ina/backfiles1.seq:517308-11	7.00	111.49	159.77	136	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-299-498A-22 -	7.00	111.49	159.77	136	
/cgn2_6/ptodata/2/ina/PCPUS_COMB.seq:PCF-US95-10813-22	7.00	101.03	169.51	145	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-564-805-26 -	7.00	101.03	169.51	145	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-090-523-39 +	7.00	108.93	221.80	194	

/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-398-627-39 + 7.00 108.93 221.80 194  
/cgn2\_6/ptodata/2/ina/backfiles1.seq:5508199-8 - 7.00 108.93 221.80 194  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-765-332-128 + 7.00 107.13 279.30 249  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-448-894-128 + 7.00 107.13 279.30 249  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-821-994-58 + 7.00 106.96 285.51 255  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-030-607-210 + 7.00 106.93 286.54 256  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-439-313-210 + 7.00 106.93 286.54 256

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-743-637B-192

seq\_documentation\_block:  
; Sequence 192, Application US/08743637B  
; Patent No. 5994066  
; GENERAL INFORMATION:  
; APPLICANT: BERGERON, Michel G.  
; APPLICANT: PICARD, Francois J.  
; APPLICANT: OUELLETTE, Marc  
; APPLICANT: ROY, Paul H.  
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
; NUMBER OF SEQUENCES: 273  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: QUARLES & BRADY  
; STREET: 411 EAST WISCONSIN AVENUE  
; CITY: MILWAUKEE  
; STATE: WISCONSIN  
; COUNTRY: USA  
; ZIP: 53202-4497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/743,637B  
; APPLICATION NUMBER: US/08/743,637B  
; FILING DATE: 04-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/526,840  
; FILING DATE: 11-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BAKER, Jean C.  
; REGISTRATION NUMBER: 35,433  
; REFERENCE/DOCKET NUMBER: 850586.90012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (414) 277-5000  
; TELEFAX: (414) 277-5591  
; INFORMATION FOR SEQ ID NO: 192:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 732 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-743-637B-192

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-647-377-9 x US-08-743-637B-192 ..  
Align seg 1/1 to: US-08-743-637B-192 from: 1 to: 732

369 ArgAtGAspLeuAtGlyAspCys 376  
|||||  
383 AGNAGGGATTTCGGAAGATTC 406

```
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-247-155-44
seq_documentation_block:
; Sequence 44, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 44
; LENGTH: 893
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 42..755
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 42..200
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.8
; OTHER INFORMATION: seq ILSQLVLTWTTS/TV
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 860..865
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 878..893
; US-09-247-155-44

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-09-247-155-44 ..
Align seg 1/1 to: US-09-247-155-44 from: 1 to: 893

221 GlyLeuSerGluLysSerLeuGln 228
|||||
137 GGCGCTTCTGAGAAAAGTCTACAG 160

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-853-1
seq_documentation_block:
; Sequence 1, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
```

```
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-911-853-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-08-911-853-1/rev ..
Align seg 1/1 to reverse of: US-08-911-853-1 from: 1 to: 1029

538 AtgArgGlnArgPheValAspGly 545
|||||
252 CGCGCCGACGGTTCGTCGATGGC 229

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-409-1
seq_documentation_block:
; Sequence 1, Application US/09479409
; Patent No. 6225106
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,409
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-479-409-1

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-09-479-409-1/rev ..
Align seg 1/1 to reverse of: US-09-479-409-1 from: 1 to: 1029

538 ArgArgGlnArgPheValAspGly 545
|||||
252 CGGCGCCAGCGGTTCGTCGATGCG 229

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-479-453-1

seq_documentation_block:
; Sequence 1, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Ouax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

US-09-479-453-1

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-09-479-453-1/rev ..
Align seg 1/1 to reverse of: US-09-479-453-1 from: 1 to: 1029

538 ArgArgGlnArgPheValAspGly 545
|||||
252 CGGCGCCAGCGGTTCGTCGATGCG 229

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-627F-22

seq_documentation_block:
; Sequence 22, Application US/08358627F
; Patent No. 6177242
; GENERAL INFORMATION:
; APPLICANT: Changeux, Jean-Pierre
; APPLICANT: Picciotto, Marina
; APPLICANT: Bessis, Alain
; TITLE OF INVENTION: Genomic DNA Fragments Containing
; TITLE OF INVENTION: Regulatory and Coding Sequences For The B2-Subunit of the
; TITLE OF INVENTION: Neuronal Nicotinic Acetylcholine Receptor and Transgenic
; TITLE OF INVENTION: Animals Made Using These Fragments or Mutated Fragments
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,627F
; FILING DATE: 14-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0135-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-358-627F-22

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-08-358-627F-22 ..
Align seg 1/1 to: US-08-358-627F-22 from: 1 to: 1289

```

```

514 IleSerProGluHisSerValGlu 521
397 ATCTCCCAAGACACAGCGTGGAA 420

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-814A-3

seq_documentation_block:
; Sequence 3, Application US/08481814A
; Patent No. 5869040
; GENERAL INFORMATION:
; APPLICANT: Hsu, Yen-Ming
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Biogen, Inc.
; STREET: 14 Cambridge Center
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02142
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.814A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaplan, Warren A
; REFERENCE/DOCKET NUMBER: A001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-679-2000
; TELEFAX: 617-679-2838
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 63..1301
; OTHER INFORMATION: /function= "transcription factor"
; OTHER INFORMATION: /product= "E2F-4"
US-08-481-814A-3

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-g x US-08-481-814A-3 rev ..

Align seg 1/1 to reverse of: US-08-481-814A-3 from: 1 to: 1332

19 valSerSerValGlyProSer 26
|||||
949 GTGTCAGGTGTGTGGCCCACT 926

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-282-305-3

seq_documentation_block:
; Sequence 3, Application US/09282305
; Patent No. 6287843
; GENERAL INFORMATION:
; APPLICANT: Baldwin, Donald A.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
; FILE REFERENCE: 5718-44,
; CURRENT APPLICATION NUMBER: US/09/282.305
; CURRENT FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/080.563
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1475
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29)..(1084)
US-09-282-305-3

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-g x US-09-282-305-3 ..

Align seg 1/1 to: US-09-282-305-3 from: 1 to: 1475

686 HistyAlaLeuAsnValProLeu 693
|||||
407 CATTATGCTCTGATGTTCCCTG 430

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-836-582-1

seq_documentation_block:
; Sequence 1, Application US/08836582
; Patent No. 6045999
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; APPLICANT: Beijersbergen, Roderick L
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye
; STREET: 1100 No. 6045999th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836.582
; FILING DATE: 14-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00868
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA: GB 9423049.7
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Arthur R. Crawford
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 620-18

```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
US-08-836-582-1

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-647-377-9 x US-08-836-582-1/rev ..
  Align seg 1/1 to reverse of: US-08-836-582-1 from: 1 to: 1489

      19 ValSerSerValValGlyProSer 26
      |||||
      913 GTGTCCAGTGTGTGGGCCAGT 890

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-265-566-1

seq_documentation_block:
; Sequence 1, Application US/09265566
; Patent No. 6303335
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye
; STREET: 1100 No. 630335th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,566
; FILING DATE: 10-Mar-1999
; CLASSIFICATION: <Unknown>
; 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,582
; FILING DATE: <Unknown>
; APPLICATION NUMBER: GB 9423049.7
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Arthur R. Crawford
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 620-18
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-265-566-1

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
US-08-836-582-1

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-647-377-9 x US-09-265-566-1/rev ..
  Align seg 1/1 to reverse of: US-09-265-566-1 from: 1 to: 1489

      19 ValSerSerValValGlyProSer 26
      |||||
      913 GTGTCCAGTGTGTGGGCCAGT 890

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-282-305-1

seq_documentation_block:
; Sequence 1, Application US/09282305
; Patent No. 6287843
; GENERAL INFORMATION:
; APPLICANT: Baldwin, Donald A.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
; FILE REFERENCE: 5718-44
; CURRENT APPLICATION NUMBER: US/09/282,305
; CURRENT FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/080,563
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1826
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29)..(1405)
US-09-282-305-1

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-647-377-9 x US-09-282-305-1 ..
  Align seg 1/1 to: US-09-282-305-1 from: 1 to: 1826

      686 HistYrAlaLeuAsnValProLeu 693
      |||||
      728 CATTATGCTCTGATGTTCCTCGT 751

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-579-777A-1

seq_documentation_block:
; Sequence 1, Application US/08579777A
; Patent No. 5912153
; GENERAL INFORMATION:
; APPLICANT: Enderlin, Carol S.
; APPLICANT: Selitrennikoff, Claude P.
; TITLE OF INVENTION: (1,3)B-glucan synthase genes and
; TITLE OF INVENTION: Inducible inhibition of fungal growth using the antisense
; TITLE OF INVENTION: constructs derived therefrom.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis, Graham & Stubbs L.L.P.
; STREET: 370 Seventeenth Street
; CITY: Denver
```

```
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80201-0185
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,777A
; FILING DATE: 28-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,004
; FILING DATE: 18-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)892-7370
; TELEFAX: (303)893-1379
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2585 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Neurospora crassa
; STRAIN: wild-type 74-DR 23-1VA
; DEVELOPMENTAL STAGE: Mycelial
; IMMEDIATE SOURCE:
; LIBRARY: lamda Zap cDNA
; CLONE: gs-1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: linkage group V
; US-08-579-777A-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-08-579-777A-1 ..
Align seg 1/1 to: US-08-579-777A-1 from: 1 to: 2585

603 AlaLeuArgHisPro 610
|||||
1458 GCGTTGCTGAGGAGGCCGCCA 1481

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-228-986-11

seq_documentation_block:
; Sequence 11, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 3097
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
; US-09-228-986-11
```

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-09-228-986-11 ..
Align seg 1/1 to: US-09-228-986-11 from: 1 to: 3097

653 LeuAlaArgLysGluValLeuThr 660
|||||
909 TTGCACGAAAGAGTCTGACG 932

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-911-853-28

seq_documentation_block:
; Sequence 28, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-911-853-28

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x US-08-911-853-28/rev ..
Align seg 1/1 to reverse of: US-08-911-853-28 from: 1 to: 4377

538 ArgArgGluArgPheValAspGly 545
|||||
```

600 CGCGCCAGCGGTCGTCGATGGC 577  
seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-479-409-28  
seq\_documentation\_block:  
; Sequence 28, Application US/09479409  
; Patent No. 6225106  
; GENERAL INFORMATION:  
; APPLICANT: Gerritse, Gijbert  
; APPLICANT: Quax, Wilhelmus J.  
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
; TITLE OF INVENTION: EXPRESSION LEVELS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genencor International  
; STREET: 925 Page Mill Road  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1013  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/479,409  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/911,853  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Glaister, Debra J  
; REGISTRATION NUMBER: 33,888  
; REFERENCE/DOCKET NUMBER: GC361-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-846-7620  
; TELEFAX: 650-845-6504  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-479-409-28

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x US-09-479-409-28/rev ..

Align seg 1/1 to reverse of: US-09-479-409-28 from: 1 to: 4377

538 ArgArgClnArgPheValAspCly 545  
|||||  
600 CGCGCCAGCGGTCGTCGATGGC 577







```

; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
; TITLE OF INVENTION: Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,165A
; FILING DATE: 24-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,032
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmì, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1348 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-936-165A-186

alignment_scores:
      Quality: 270.50      Length: 350
      Ratio: 1.346        Gaps: 13
      Percent Similarity: 57.429      Percent Identity: 26.286

alignment_block:
US-09-647-377-9 x US-08-936-165A-186      ..

Align seg 1/1 to: US-08-936-165A-186 from: 1 to: 1348

446 LeuCySGluGluLeuCySSerLeuAsnProMetThrAspLysLeuThrPh 462
||||| : : : : : ||||||| : : : : : ||| : : : : :
8 CTAGCAATGGTATTGTTTCATTAACTCCTATGTTGATCGTTTAAAGTT 57

462 eSerValIleTrpLysLeuThrProGluGlyLysIleLeuGluGluTrpP 479
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 AAGCTGTCGATCGAATCGATGCTAGTGGTCGCTTGTATAACATGAAA 107

479 heGlyAArgThrIleAArgSerCysThrLysLeuSerTyrAspHisAla 495
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
108 TTTNTGATAGTTATACATCTGATTATTCGAATCACGTATGATGCGGTA 157

496 GlnSerMetIleGluAsnProThrGluLysIlePro.....GluGluG 510
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
158 AATCAGATTATT.....ACTGAAAGGATCCTACATCGCGGAA 198

510 uLeuProPheSerProGluHisSerValGluGluValHisGlnAlav 527
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
199 ATATAAGAAATATACCGCTATGTTAGATTATGACACAA..... 235

```

```
seq_name: /cyn2_6/ptodata/2/1na/6B_COMB.seq:US-08-936-165A-186
seq_documentation_block:
; Sequence 186, Application US/089361615A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
```

```

527 aLeuAsnLeuHisSerIleAlaIysGlnLeuArgArgGlnArgPheVal 543
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
236 ....GATTTCATCTAATCGTTTGAATCAATGAGAAACAGCGTGGGTG 280
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
544 AspGlyAlaLeuArgLeuAspGlnLeuIleLeuAlaPheThrLeuAspHi 560
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
281 AANTCGATTGTGATATATTAGNAGCAAGAAAGTA.....TTAGTTAA 321
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
560 sGlu...ThrGlyLeuProGlnGlyCysHisIleTyrArgAspSer 576
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
322 CGAAGTCGGGTATACCAACAGATGTTCAATTAAGACAACAGCGCGGGT 371
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
577 AsnIysLeuValGlnGluPheMetLeuLeuAlaAsnMetAlaValAlaHi 593
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
372 GAACGCTCAATTGAATCATTTATGTTAATGCAATGAACAGTGTGCTGA 421
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
593 sIysIlePheArgThrPheProGluGlnAlaLeuLeuArgHisProp 610
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
422 ACAT...TTTAGTAAGATTAAATGTACCTTTTATTACCGAGTCATGAGC 468
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
610 roProGlnThrLysMetLeuSerAspLeuValGluPheCysAspGlnMet 626
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
469 AACCTAATCAGATCCCTTAAGACAATTCCTGATTATTATACAAACTTT 518
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
627 GlyLeuProMetAspValSerSerAlaGlyAlaLeuAsnLysSerLeuTh 643
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
519 GGCATC.....ATGAT 529
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
643 rLysThrPheGlyAspAspLysTyrSerLeuAlaArgLysGluValLeuTh 660
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
530 TAAGGCATCGCGGAAGATATTCATCCACACACATTCRAAAGGTTCAAG 579
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
660 hrAsnMetTyrSerArgProMetGlnMetAlaLeuTyrPheCysSerGly 676
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
580 AAGAAGTAGAAGTCGACCTGAACAAATGTGCTATT.....TCAACAATG 623
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
677 MetLeuGlnAspGlnGluGln.....PheArgHi 686
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
624 ATGTTAGTTCATCGACACAGCGCATTTATGATGATGAACCTTGGGACA 673
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
686 sTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIleArgA 703
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
674 TTGGGCTTATCAGCTGAATATATAGCATTTNACATCACAATTAGAC 723
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
703 rgPheAlaAspValIleValHisArgLeuAlaAlaLeuGlyTyr 719
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
724 GTTATCCTCATTTACAGTCATCGTTTAAATCCGTAGATATTAATGTAG 773
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
720 SerGluGlnProAspValGlu.....ProAspThrLeuGlnIly 732
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
774 AATTCATCGGATCAACAAAGAGTGAAGCGTTGGGAAGACAATTCGCTGA 823
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
732 sGlnAlaAspHisCysAsnAspArgArgMet.....AlaSerL 745
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
824 GTTAGCTGAACATCTCTTAACGCTGAACGCTGCTGCTATTGGAGCAGAAC 873
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
745 ysArgValGlnGlnLeuSerIleGlyLeuPhePheAlaValLeuVallys 761
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
874 GTGATACTCATGAATTCGAAAAAGCAGAAATAT.....ATGATTCAA 914
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
762 GluSerGlyProLeuGlnSerGlnAlaMetValMetGlyValLeuAsn 777
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
915 CATATTGGTGCAT...GAATTTGAAGGTATTTGCAGCTACGTAGCTAAC 959
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: /cyn2_6/ptodata/2/ina/6B_COMB_seq:US-09-327-487A-2

```

```
seq_documentation_block:
; Sequence 2, Application US/09327487A
; Patent No. 6352847
; GENERAL INFORMATION:
; APPLICANT: MATSUKAWA, Hirokazu
; APPLICANT: OKA, Osamu
; APPLICANT: FUJITA, Tuvosi
```

```

: APPLICANT: MIYAZAKI, Kentaro
: TITLE OF INVENTION: AMMONIA ELIMINATION LIQUID REAGENT
: FILE REFERENCE: 028022-013
: CURRENT APPLICATION NUMBER: US/09/327.487A
: CURRENT FILING DATE: 1999-06-08
: PRIOR APPLICATION NUMBER: JP 10-176643
: PRIOR FILING DATE: 1998-06-09
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO. 2
: LENGTH: 2872
: TYPE: DNA
: ORGANISM: Thermus aquaticus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (180)..(1460)
US-09-327-487A-2

alignment_scores:
    Quality: 235.00      Length: 513
    Ratio: 0.955        Gaps: 24
    Percent Similarity: 47.953    Percent Identity: 24.951

alignment_block:
US-09-647-377-9 x US-09-327-487A-2/rev ..

Align seg 1/1 to reverse of: US-09-327-487A-2 from: 1 to: 2872

350 LeuGluCysLeuProGlnSerLeuProThrThrLeProProAspGluVa 366
      ||||| :|||||
2679 CTTCCCTCGCCCGCGGACGCTCCCC.....GTCGCCCTTCGCCGA 2636
      :|||||
366 lGlyLysArgArgAspLeuArgLysAspCysIlePheThrIleAspProS 383
      :|||||
2635 GGAGGAGGGGTGGACCTCACCCACCTCCCGCCTTCGCCATTGACGAGC 2586
      :|||||
383 erThrAlaArgAspLeuAspAlaLeuAlaCysArgLeuThrAsp 399
      :|||||
2585 AGGGGAGCCAGGACCCGGACGAGCCCTTACGCCGAAGGGTGGAGGAG 2536
      :|||||
400 GlyThrPheGluValGlyValHisIleAlaAspValSerTyrPheValPr 416
      :|||||
2535 GGC...TTCACCTTTTCGTCATGTGGCGGATGTGGCCGCTTGTGTCGG 2489
      :|||||
416 oGluGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValT 433
      :|||||
2488 GCGGGAGAGCCCTCGACACGAGGAGCCCTCCGCCGGGGGCCAACCTCT 2439
      :|||||
433 yrluValGlnLysValValPrometLeuProArgLeuLeuCysGluGlu 449
      :|||||
2438 ATCTGCCGAGGGCAGCGGTGCCATGCTCCCGCTCCCGGTGACCGAGGCC 2389
      :|||||
450 Leu...CysSerLeuAsnProMetThrAspLysLeuThrPheSerValIle 465
      :|||||
2388 CTGGGCTGGGCTTAAGGAGGTCTCCCGAGCCCTCACCTTTGAGCTTCG 2339
      :|||||
466 TtpLysLeuThrProGluGlyLysIleLeuGluThrPheClyArgTh 482
      :|||||
2338 .....GTCCTCCCGAGGA.....GCCTTTGGGGGAGGAC 2307
      :|||||
482 rIleIleArg...SerCysThrLysLeuSerTyrAspHisAlaGlnSerM 498
      :|||||
2306 CTTTACCTCTCCTGGGTGGGTGAGGGCGGCTT..... 2273
      :|||||
498 etIleGluAsnProThrGluLysIleLeProGluGluLeuLeuProProl 514
      :|||||
2272 .....GTCCTCCCGAGGA.....CTACCG..... 2267
      :|||||
515 SerProGluHisSerValGluGluValHisGlnAlaValLeuAsnLeuHi 531
      :|||||
2266 .....GAGGCTTCGAGGTG...GAGGCCCTGGCGCCATGAA 2232
      :|||||

```



```

649 ASPLYSYrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerAr 665
    ||| :|||
448 TTAAATCAGATTAGCA..... 465
665 gPrometGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnG 682
465 ..... 465
682 luGlnPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThr 698
    ||| :|||
466 .....CCGCACCTTGGTCTTGAAGGCTATGCCACTTGGACA 507
699 SerProIleArgPheAlaAspValIleValHisArgLeuLeuAla 715
    ||| :|||
508 TGCCCATCCGCAATATTCAGATATGTTAATCATCGCTTAATCAAGC 557
715 aaLeuLeuGlyTyrSerGluGlnProAspValGluPro.....AspThrL 730
    ||| :|||
558 CGTGTG.....GCNAACAGCTTATGAACCAACCAAAATGACGTGT 601
730 euGlnLysGlnAlaAspHisCysAsnAspArgArgMetAlaSerLysArg 746
    ||| :|||
602 TGCCACGCTTCCGAGAGCTCGCCGCCGCAAAATCGCTAGTGAACGTAT 651
747 ValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuValLysGluSe 763
    ||| :|||
652 ATTGGCGATGGCTATATTCGCTTATCTGTGACAAAGTGGCTGAAA 701
763 rGlyProLeuGlnSerGluAlaMetValMetGlyValLeuAsnGlnAlap 780
    ||| :|||
702 T.....GTGGAATTTAATCAGAAGTGAAGATGTAATGCGTCAAGCT 745
780 heAspValLeuLeuArgPheGlyValGlnLysArgIleTyrCysAsn 796
    ||| :|||
746 TACGCGTACAACTGCTCGAAATGNGCATCG.....CTATTATTCCT 789
797 AlaLeuAlaLeuArgSerTyrSerPheGlnLysValGlyLysLysProGl 813
    ||| :|||
790 GCCGCCAGTTGCACAAC.....AACAAAGA 818
813 uLeuThrLeuValTrpGluProAspLeu.....GluGluG 826
    ||| :|||
819 AATACAGCTA.....AACCTGACGAACTGCCCTCTATATAAAGCG 862
826 luProThrGlnGlnValIleThrIlePhe..... 835
863 AACGCACTTACAAATAGCGGACATTGTGAAGTGAACCTACAGAAAGTG 912
836 .....SerLeuValAspValValLeuGlnAlaGluAlaLath 847
    ||| :|||
913 AAAGAAGCACTCGCAGTATTGTGGCGAAATACTTCAATAAATGCGCT 962
847 tAlaLeuLysTyr 851
963 TCCAATATGTAC 975

```

seq\_name: /cgn2.6/tbdata/2/ina/5A\_COMB.seq:us-08-257-073-10

seq\_documentation\_block:

; Sequence 10, Application us/08257073

; Patent No. 5766597

; GENERAL INFORMATION:

; APPLICANT: Paolletti, Enzo

; APPLICANT: de Taisne, Charles

; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.

; STREET: 530 Fifth Avenue, 25th Floor

; CITY: New York

; STATE: New York

; COUNTRY: UNITED STATES OF AMERICA

```

;
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-10

```

alignment\_scores:

Quality: 124.50 Length: 902

Ratio: 0.310 Gaps: 43

Percent Similarity: 44.457 Percent Identity: 19.401

alignment\_block:

US-09-647-377-9 x US-08-257-073-10 ..

Align seq 1/1 to: US-08-257-073-10 from: 1 to: 5181

```

75 ileAsnPro.....LysLysPheHisGluAlaPheIleProSerPr 88
    :|||
1387 ATGAATCTGTACTAAAGAAAAAATTAAGAAAAAATATT...ACAGA 1433
    :|||
88 oAspGlyAspArgAspIlePheIleAspGlyValValAlaArg..... 102
    :|||
1434 TAATAAGGAAAGAAAAATATTATTATTAACATTAAAAAACAATTTGATT 1483
    :|||
103 .....AsnArgAlaLeuAsnGlyAspLeuValValLysLeuPro 117
    :|||
1484 TAGAAGAAAAAATTAATTAATTAATTTGACAAAGATGCTGATAGATAAAA 1533
    :|||
118 GluAspGlnTrpLysAlaValLysProGluSerAsn..... 129
    :|||
1534 GAAGATTATGAAAGTCAAAAAGGATTATGAAGAATTTACTTGAAAAATT 1583
    :|||
130 .....AspLysG 132
    :|||
1584 TTATGAATGAATTAATTAATTAATTTGACAAAGATGCTGATAGATAAAA 1633
    :|||
132 IuIleGluAlaThrTyrGluAlaAspIleProGluGluGlyCysGlyHis 148
    :|||
1634 TATTCAAGTCAAGATATACATATAATATTGAAACAAAGATATAATAAT 1683
    :|||
149 .....HisProLeuGlnGlnSerArgLy 156
    :|||
1684 AAATTTTCATCTCTAATAATTTCTGTATATATATGTTCAAAANTTAAAAA 1733
    :|||

```



```

1      |||
3288 ACAATTAGAAATCAAAATTTGAATCCTTAATAACCCACATATGTATTAC 3337
671 euTyrPheCysSerGlyMetLeuGlnAspGlnGluGln..... 683
3338 AAAACTTTTCGTGTTTCTTTAAACAAAAAAGAAAGACTGAATAACAGCAA 3387
684 .....PheArgHisTyrAlaLe 689
3388 ACTGAAACACATTAGAAAAACAAAAATATTATTGAACATTATAAAGG 3437
689 uasnValProLeuTyrThrHisPheThrSerProIleArgArgPheAla 706
3438 ACTGTTAAATATTATTAATGGTGAATCATCTCCATTAAAAACTTTAAGTG 3487
706 spValIleValHisArgLeuLeuAlaAlaLeuGlyTyrSerGluGln 722
3488 AAGTATCAATTCAACA.....GAAGATAATTAT 3516
723 ProAspValGluPro...AspThrLeuGlnLysGlnAlaAspHisCysAs 738
3517 GCCAATTTAGAAAAATTTAGAGTATTAGTAAATAGATGGAACCTCAA 3566
738 nAspArg...ArgMetAlaSerLysArgValGlnGlnLeuSerIleGlyL 754
3567 TGATAATTACATTTAGAAAGAAAAAATATCTTCTATCAATGAT 3616
754 euPhePheAlaValLeu...ValLysGlu..... 762
3617 TACATCATTTAATTACTGAATTAAGAAGTAATAAATAAATAAATTAT 3666
763 .....SerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAs 777
3667 ACAGTAATTCCTCAAGTGAATAATRAAGAAAGTT.....AA 3704
777 nGlnAlaPheAspValLeuValLeuArgPheGlyValGlnLysArgIleT 794
3705 CGAAGCTTAA.....ATCTAGGAAATTTCTCCAGGAAGCAAAAGTT 3750
794 yr.....CysAsnAlaLeuAlaLeuArgSerTyrSerPheGlnLysVal 808
3751 ACACAGTTGTAACCTCAACCTCAACACAGATGATACTCATCTCATTATC 3800
809 GlyLys 810
3801 TGTAAG 3806

seq_name: /cnp2_6/ptodata/2/ina/6A_COMB.seq:us-09-335-409-1

seq_documentation_block:
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

```

alignment\_scores:

```

Quality: 121.50      Length: 1023
Ratio: 0.292        Gaps: 46
Percent Similarity: 40.665      Percent Identity: 18.084

alignment_block:
US-09-647-377-9 x US-09-335-409-1

Align seg 1/1 to: US-09-335-409-1 from: 1 to: 68750

11 ArgSerProGlyThrProArg.....GlyValSerSerVa 22
||||: ||| |||
60857 AGGCGCAGGCGCTGCGCGCTGAGCATGAGTGGCGCATGTTCAACGA 60906

22 lValGlyProSerAlaValGlyAlaSerProGlyAspLysLysSerLysA 39
||||| :|||: |||
60907 GGTGGGATGGCGGTGCGCAAGAAACCGTGGCGCGGAGATC...T 60953

39 snLysSerMetArgGlyLysLysSerIlePheGluThrTyrMetSer 55
:||||: ||| |||
60954 CTGCGGGATGCGGGC.....ATCAC 60976

56 LysGluAspValSerGluGlyLeuLysArg.....GlyThrLeuI 69
||||: :|||: |||
60977 CCCGATGAGGCTCTGCTAGCTCTGCGCGCTTCTCGAGGGTATCGGT 61026

69 eGln...GlyValLeuArgIleAsnProLysLysPheHisGluAlaPheI 85
:||||: |||: |||
61027 GCAGCGGGGTGATACCGATCCTCGCGGAGTGGTGAG...TTCT 61073

85 leProSer..... 87
||||:

61074 ACCCGCAACAGCGCCTCAGGAGTTGTCGGGCTGTGACCAACGAG 61123

88 .....ProAspGlyAspArgAspIlePheIleAspGI 98
||||: |||: |||
61124 CGCGGCTCGCTATCGGACCGCGGGATCGGACCTGCTCGAACAGCT 61173

98 yValValAlaArgAsnArgAlaLeuAsnGly.....AspLeuValV 112
:||||: |||
61174 TGGCTGGGTGAGCGCGCGGGGGGTGCTGCGAGGACGCTGCTGC 61223

112 alValLysLeu.....LeuProGluAspLysTrpLysAla 123
||||: |||
61224 CGTGCAGGTCTCGCATGTGCTGGCTCTCCCTGAAGACAAGATCGAG 61273

124 ValLysProGluSerAsn..... 129
||| |||:

61274 GATGCGCGCTCTCGAGCATGGCATGCTCGCTGATGACCTGGAGCT 61323

130 .AspLysGluIleGluAlaThrTrpGluAlaAspIleProGluGluGlyC 146
:||||: |||: |||
61324 GCGCAACGATCGAGGCTGCTGGGCGTGGCGGCTGCTCGACGAC 61373

146 ysGlyHisHisProLeuGlnGlnSerArgLysGlyTrpSerGlyProAsp 162
:||||: |||
61374 GTTGACGTACCAACGATAGCAGCATACCGCTGCTCTCGACGAC 61423

163 ValIleIleGluAlaGlnPheAspAspSerAspSerGluAspArgHisGI 179
:||||: |||: |||
61424 GCCTCGTGGTGGCGGCTGGCGGCTGCGACACGAGCAATCGACGGC 61473

179 yAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerThrP 196
:||||: |||
61474 GAGCGCGGTTCTGCTGCTCCACGCTCTCCG..... 61504

196 roAspArgGlyLysGluAspSerSerThrProValMetLysAspGlu... 211
||||: |||: |||
61505 .....TTTCGTCTCTGCTGAAGCGCGGCT 61531

212 .....AsnThrProIleProGlnAspThrArgGI 221
||||: |||
61532 CGTCTCTTCTGTTTACGGGCTTCTGGCGGCTGCGCGAGGGCTTCGGTTC 61581

```

221 yLeuSerGluLysSerLeuGlnLysSerAlaLysValValTyrIleLeuG 238  
: ||||| ||||| : : : : : : : : : : : :  
61582 CTGGTCGGAAGATCTCAGTGGACCGACTGTGAAATCTGTGCCCATGTGC 61631

238 LuLySylHisSerArgAlaAlaThrGlyIleLeuLysLeuLeuAlaAsp 254  
: : : : : : : : : : : : : : : : : :  
61632 ACATCCGAGCTCCCTCCGAGGACGCCCTGTGTAAGAAGTAGCTCCAA 61681

255 LysAasNerAsPleuPhelLysLysTyrAlaLeuPheSerProSerAspHi 271  
: : : : : : : : : : : : : : : : : :  
61682 GAGCGCGCTCGCTGATTCACTATGCA.....TTAGTA 61731

271 sArgValProArgIleTyrValProLeuLysAspCysProGlnAspHem 288  
: : : : : : : : : : : : : : : : : :  
61712 .....GAGCGACCG..... 61720

288 etThrArgProLysAspPheAlaasnThrLeuPheIleCysArgIleIle 304  
: : : : : : : : : : : : : : : : : :  
61721 .....TTTTGG.....TTAGTA 61731

305 AspTrpLysGluAspCysAsnphelaLeuCluy..... 315  
: : : : : : : : : : : : : : : : : :  
61733 GGGTTcAGCGTGCGTTCGTTcGTATGGGACAGCGGTGAGTCGC 61782

316 .....GlnLeuAlaLysS 320  
: : : : : : : : : : : : : : : : : :  
61783 CAGTCGTTCCGGCGCACCGGCTCGGTCGTCACGTTGGCGGCGCA 61832

320 erLeuGlyGlnAlaGlyGluIleGluProGlnThrGluGlyIleLeuThr 336  
: : : : : : : : : : : : : : : : : :  
61833 GCRTGATCTCTTCACAGATCATCCCCGAGATGGAGACCGATATA... 61879

337 GluTyrGlyValAspPheSerAspPhe.SerSerGluVal.....LeuG 351  
: : : : : : : : : : : : : : : : : :  
61880 .....ATAGCCAAGCTCTTCTCCGAATGCCCGGGTTT 61914

351 lucYsLeuPro..... 354  
: : : : : : : : : : : : : : : : : :  
61915 CGTGGATCCACCACAAGTCCAGCGGATGTCGCGCAGACAAGGTCA 61964

355 GlnSerLeuProThrIleProProAspGluValGlyLysArgArgAs 371  
: : : : : : : : : : : : : : : : : :  
61965 TCACAGACACCAATGGTGGCTCGGCCCGGAGCTCGAAGAGCGCGCCC 62014

371 pLeuArgLysASPCySillePheThrIleAspProSerThrAlaLaargSpL 388  
: : : : : : : : : : : : : : : : : :  
62015 GTGAAGATCGCGTCCCTATCGTG.....CCATCGCGCGGCTCGAGCA 62058

388 euAspAspAlaLeuAlaCysArgArgLeuThrAspGlyThrPheGluVal 404  
: : : : : : : : : : : : : : : : : :  
62059 TG.....TCATCGTCCCTCCGAGCGAGCTTCAGGATCTACAAATCTC 62099

405 GlyValHistleAlaAspValserTyrPheValProGluLysSerLe 421  
: : : : : : : : : : : : : : : : : :  
62100 GCACACCGGAGCGCTTCTATATGATCTCTTCCTCCGAGATCACGAAT 62149

421 uAspLysValAlaAlaGluArgAlaThrSerValTyrLeuValGlnLysV 438  
: : : : : : : : : : : : : : : : : :  
62150 CTCGTGATCGAGGCGCGAGATCATGCATC.....G 62183

438 alValProMetLeuProArgLeuLeuCysGluGluLeuCysSerLeuAs 454  
: : : : : : : : : : : : : : : : : :  
62184 TCGACTCGCATCTCAATCGCGTGCWCGCGGAGAGAGCGTCTCGAGC 62233

455 ProMetThrAspLysLeuThrPheSerValIleThrLysLeuThrProGl 471  
: : : : : : : : : : : : : : : : : :  
62234 CCCGCGTTCAG.....GCAAAATGATGGGAG.....CTCC 62265

471 uGlyLysIleLeuGlnGluThrPheGlyArg..... 481  
: : : : : : : : : : : : : : : : : :  
62266 CTCGGGCGCGAGATGGTTGGGAGACACGCTGGCGCTGGCGCGCGCG 62315

482 .....ThrIleIleArgSerCysThrLysLeuSer 491

62316 CAGCCGCGGAGCCCATGACCTTCTTGACGGTTGCAGT.....ATA 62359  
::: ::::: |||:::  
492 TyrAspHisAlaGlnSerMetIleGluAsnProThrGlyLysIleProGl 508  
||| :: |:: :||| :|||  
62360 GGAGATTTTATCACACAGGAGCAACGAATTCAGAGTGTGACCAAGCCTGC 62409  
::: ::::: |||:::  
508 uGluGluLeuProIleSerProGluHis.....S 519  
::: ::::: |||:::  
62410 TTTCGACTTCAGCGCTTCGGCGCTGGGTACGCGAGGACCGTTCGCCG 62459  
::: ::::: |||:::  
519 erValGluGluValHisGlnLa..... 526  
::: |||::: |||::: |||  
62460 CGATCGAGCGCTGTGAGAGGACCACCCCATTCTTCTACGGGTGATGAGGC 62509  
::: ::::: |||:::  
527 .....ValLeuAsnLeuHisSerIleAlaLysGlnLeuArgAr 539  
::: ::::: |||:::  
62510 CGCTCTGGGTCTCTACCCCGATACACGAGTGTGCGCGTGTTCGCGA 62559  
::: ::::: |||:::  
539 gLnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlap 556  
::: |||::: |||::: |||  
62560 GGAAGCGTTCGCG.....GTACGTGAGAAGAGTGGGAAT 62594  
::: ::::: |||:::  
556 heThrLeuAspHisGluThrGlyLeuProInGlyCystHisIleTyrGlu 572  
::: ::::: |||:::  
62595 CGAGCGGGAGTAGTACTCGTCGGCCATTCCCGAG..... 62626  
::: ::::: |||:::  
573 TyrArgAspSerAsnLysLeuValGluLpheMetLeuLeuAlaAsnMe 589  
::: ::::: |||:::  
62626 ..... 62626  
589 tAlaValAlaHisLysIlePheArgThrPheProGluGlnAlaLeuLeuA 606  
::: ::::: |||:::  
62626 ..... 62626  
606 rGArgHisProProGlnThrLysMetLeuSerAspLeuValGluPhe 622  
|||||::: |||::: |||:::  
62627 .....CTCAGCATATGAAGAAGTAC 62647  
::: ::::: |||:::  
623 CysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeuAs 639  
::: |||::: |||::: |||  
62648 ..GSATTGTTCGGGTGCGCGGAG.....GATCAGCTCGGGTCCG 62688  
::: ::::: |||:::  
639 nLysSerLeuThrLysThrPheGlyAspAspLysTyrSerLeuAlaArgL 656  
|||::: |||::: |||::: |||  
62689 CAAGCTCGTCAACCCGTCGTTTACGTACACGCGCATCGACCTGTCGCGC 62738  
|||::: |||::: |||::: |||  
656 ysGluVaLeuThrAsnMetTyrSerArgProMetGlnMetAlaLeuTyr 672  
|||::: |||::: |||::: |||  
62739 CCGAAATA.....CAGCGCACCGTCTGACACAGTCTCGAT 62773  
|||::: |||::: |||::: |||  
673 PhecysSerGlyMetLeuGlnAspGlnGluPhe.....ArgHis 686  
|||::: |||::: |||::: |||  
62774 GCTCGCTCCGGA.....CAAGAGGAGTTCGACGTTGTGCGGA 62811  
|||::: |||::: |||::: |||  
686 sTyrrAlaLeuAsnVaProLeuTyr..... 694  
|||::: |||::: |||::: |||  
62812 TTACCGGAGGGAATCCCGATGGCGCGCATCAGCGCTCTGTTCAAGGTTC 62861  
|||::: |||::: |||::: |||  
695 ..ThrHisPheThrSerProIleArgArgPheAlaAspValIleValHis 710  
::: ::::: |||:::  
62862 CGGCGGAGTGTGACGAGAAGTTCGTCGCTCGGCTCGGG..... 62902  
::: |||::: |||::: |||  
711 ArgLeuLeuAlaAlaLeuTyrSerGluGlnProAspValGluPr 727  
|||::: |||::: |||::: |||  
62903 ....ACTGCGCGCGCTHCGCGTGGGTTGGTGGCCCGACGTCATGA 62946  
|||::: |||::: |||::: |||  
727 oAspThr..... 729  
::: |||  
62947 GGAGACCAAGACCTCTGCGTCCGTACCGAGAGGGCTCGCGTCTCTCC 62996  
::: |||  
729 ..... 729

```
62997 ATGACGTCCTCGATGAGCGCGCAGGAACCCGCTCGAAATGACGCTCTTG 63046
730 .....LeuGlnLysGlnAlaAspHisCysAsnAspArgMetal 743
63047 ACCATCGCTCTTCAGCGCGAGCGCAC.....GGCAGCAGCGCTGAG 63087
743 aSerLysArgValGlnGlnLeuSerIleGly..... 753
63088 CAGGAAGAGCTGGTTCGCGCTC...GTGGGTGGATTATCGCTGCTGGCA 63134
754 .....LeuphePheAlaVal.....LeuValLys 761
63135 CCGATACACAGCTATACCTTATCGCTTCGCTGCTCAACCTGCTCGG 63184
762 GluSerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnG1 778
63185 TCGCCCGAGGCGCTCGAGCTGGTGAAGCGGAGCGCGGCTCATGAGGA 63234
778 nAlaPheAspValLeuVal.....L 785
63235 CCGCTCGATGAGTGTCTCGCTTCGACATATCTCAGATAGAACTG 63284
785 euArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArg 801
63285 TCGCTTTCGCGAGCAGGAGCTGGAGTACTCGCGGCGCATCGATC..... 63328
802 SerTyrSerPheGlnLysValGlyLysProGluLeuThrLeuValTr 818
63329 .....ASGAAGGGGAGATGGTCTTCCT 63354
818 pGluProAspLeuGluGluProThrGlnGlnValIleThrIleP 835
63355 GATCCGAGCGCCCTGAGATGG.....ACTGTAT 63386
835 heSerLeuValAspVal 840
63387 TCTCCAGCCAGAGCTG 63403

seq_name: /cgn2_5/ptodata/2/ina/6B_COMB.seq:us-09-568-102-1

seq_documentation_block:
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
alignment_scores:
  Quality: 121.50      Length: 1023
  Ratio: 0.292        Gaps: 46
  Percent similarity: 40.665  Percent Identity: 18.084
```

```
alignment_block:
US-09-647-377-9 x US-09-568-102-1
```

```
Align seg 1/1 to: US-09-568-102-1 from: 1 to: 68750
11 ArgSerProGlyThrProArg.....GlyValSerSerVa 22
60857 AGGCGCAGCGGCTCGCGCGCTGAGCATCGGCGCATGTCACGGA 60906
22 lValGlyProSerAlaValGlyAlaSerProGlyAspLysLysSerLysA 39
60907 GGTGGGATGGCGTTGCCAAGAAACCGTGGCGCGCGCAGATC...T 60953
39 snLysSerMetArgGlyLysLysLysSerIlePheGluThrTyrMetSer 55
60954 CTCGGGATGCGGGC.....ATCACC 60976
56 LysGluAspValSerClucllyLeuLysArg.....GlyThrLeuI1 69
60977 CCGATGAGGCTGTGACGCTCTGGCGGCTGCTCGAGGCTGATCGGCT 61026
69 eGln...GlyValLeuArgIleAsnProLysLysPheHisGluAlaPheI 85
61027 GCAGCGGGGCTGATACCGATCACTCCGCGCAGTGGGTGAG...TTCT 61073
85 leProser..... 87
61074 ACCCGCACACGGGCTCACGGAGTGTGCGGCTGTGACACGCGAG 61123
88 .....ProAspGlyAspArgAspIlePheIleAspG1 98
61124 CCGCGCTGCTGATCGGACCGCGGATCGGACCTGCTCGAACACGCT 61173
98 yValValAlaArgAsnArgAlaLeuAsnGly.....AspLeuValV 112
61174 TGGTGGCTGAGCCGCGCGGCGGCTGCTGCGAGGACGTGCTGTC 61223
112 alValLysLeu.....LeuProGluAspGlnTrpLysAla 123
61224 GCSTGAGGCTCGCATGTGCTCTCCTCGAAGACAAGATCGAGTGTG 61273
124 ValLysProGluSerAsn..... 129
61274 GATGCCCGCTCTCGAGCATGGGATCGGATCGCTGATGAGCTGGAGCT 61323
130 AspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGluGlyC 146
61324 GCGCAACCGCATCGAGGCTGCGTGGCGCTGCGCGCTCGAGCCTGG 61373
146 ysglyHisHisProLeuGlnSerArgLysGlyTrpSerGlyProAsp 162
61374 GGTGGACGTACCCACGCTACGACGATACCGGCTGCTGCTCGACGAC 61423
163 ValIleIleGluAlaGlnPheAspAspSerAspSerGluAspArgHisG1 179
61424 GCCTCGTCTCGCGCTTGGCGGCTGGACAGGACGGAATCGACGGC 61473
179 yAsnThrSerGlyLeuValAspGlyValLysLysLysLeuSerIleSerThrp 196
61474 GAGCGCGGTTCGTCTGCCACGCTCTCCG..... 61504
196 roAspArgGlyLysGluAspSerSerThrProValMetLysAspGlu... 211
61505 .....TTTCGCTCTGCTCGTCAAGCGCGGGCT 61531
212 .....AsnThrProIleProGlnAspThrArgG1 221
61532 CGTCTCTTCTGTTTTACGGTCTTGGCGGCTCGCCCGAGGCTTCCTTC 61581
221 yLeuSerGluLysSerLeuGlnLysSerAlaLysValValTyrIleLeuG 238
61582 CTGGTGGAGAGTCTGAGTGGAGCATCTGGAATCTGGCCATGTGGC 61631
238 luLysLysHisSerArgAlaAlaThrGlyIleLeuLysLeuLeuAlaAsp 254
61632 ACGATCGCAGCCTCGCTCGAGGAGCGCGCTGGTAAGAGTACGTCCAA 61681
```



```

63088 CAGAAAGAGCTGGTCGGCTC...GTGGGTGGATTATTCGCTGCTGCA 63134
754 .....LeupheAlaVal.....LeuValLys 761
63135 CCGATACACGATACCTATCGCTGCTCACTCACTGCTCGG 63184
762 GluSerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnG1 778
63185 TCGCCGAGGGCTCGAGCTGGTGAAGCGGAGCGGCTCATGAGAA 63234
778 nAlaPheAspValLeuVal.....L 785
63235 CCGGCTCGATGAGGTGCTCGCTCGACATATCTCAGATAGAACTG 63284
785 euArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArg 801
63285 TCGGTTTCGCGAGCGCTCGAGTACTGGAGTACTGGGGCATCGATC..... 63328
802 SerTyrSerPheGlnLysValGlyLysProGluLeuThrLeuValTr 818
63329 .....AAGAAGGGGAGATGGTCTTCTCCT 63354
818 pGluProAspLeuGluGluProThrGlnGlnValIleThrIleP 835
63355 GATCCGAGCGCCCTGAGATGGG.....ACTGTAT 63386
835 heSerLeuValAspVal 840
63387 TCTCCAGCCAGAGCTG 63403

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-567-969-1

seq_documentation_block:
; Sequence 1, Application US/09567/969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1

```

```

alignment_scores:
  Quality: 121.50      Length: 1023
  Ratio: 0.292        Gaps: 46
  Percent Similarity: 40.665      Percent Identity: 18.084

```

```
alignment_block:
```

```
US-09-647-377-9 x US-09-567-969-1 ..
```

```
Align seg 1/1 to: US-09-567-969-1 from: 1 to: 68750
```

```
11 ArgSerProGlyThrProArg.....GlyValSerSerVa 22
```

```
60857 AGGGCGAGGGCGCTGCGCGCGTGGAGCTGAGCTGGGGCATGTTTCACGA 60906
```

```
22 lValGlyProSerAlaValGlyAlaSerProGlyLysSerLysA 39
```

```

60907 GGTGGGATGGCGTTGGCAAGAAACCGTGGCGCGCGCAGATC...T 60953
39 snLysSerMetArgGlyLysLysLysSerIlePheGluThrTyrMetSer 55
60954 CTCGCGGATGCGGGC.....ATCAC 60976
56 LysGluAspValSerGluGlyLeuLysArg.....GlyThrLeuI1 69
60977 CCGATGAGGGTCTGTACGTCTGCGCGCTGCTCGAGGGTGCATCGGT 61026
69 eGln...GlyValLeuArgIleAsnProLysLysPheHisGluAlaPheI 85
61027 GCAGCGGGGGTGTATACCGATCACTCCGCGCAGTGGGTGGAG...TTCT 61073
85 leProSer..... 87
61074 ACCGGCACACGGGCTCACGGAGTTGTGCGGCTGGTGACACGCGAG 61123
88 .....ProAspGlyAspArgAspIlePheIleAspG1 98
61124 CGCGCGGTGCTGATCGGACCGCGGGATCGGACCTGCTCGAACAGCT 61173
98 YValValAlaArgAsnArgAlaLeuAsnGly.....AspLeuValV 112
61174 TGGTCTGCTGAGCCGAGCGCGGGGCTGCTGCGAGGACGTCTGTC 61223
112 alValLysLeu.....LeuProGluAspGlnTrpLysAla 123
61224 GCSTGAGGTCTCGATGTGCTGCTCTCCCTGAAGACAAGATCGAGTG 61273
124 ValLysProGluSerAsn..... 129
61274 GATGCCCGCTCTCGAGCATGGGATGGACTCGCTGATGAGCTGGAGCT 61323
130 AspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGluGlyC 146
61324 GCGCAACCGCATCGAGGTGCTGCGTGGCGCTCGCGCGCTCGAGCCTGG 61373
146 ySGlyHisHisProLeuGlnGlnSerArgLysGlyTyrSerGlyProAsp 162
61374 GTGGAGCTACCAACGGTAGCAGCATACGCGTGGCTGCTCGACGAC 61423
163 ValIleIleGluAlaGlnPheAspAspSerAspSerGluAspArgHisG1 179
61424 GCCTCTGCTCCGGCTGGCGCGGCTCGGACACGAGCAATCGACGGC 61473
179 YAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerThrP 196
61474 GAGCGCGGTGCTGCTCCACGTCTCTCCG..... 61504
196 roAspArgGlyLysGluAspSerSerThrProValMetLysAspGlu... 211
61505 .....TTTCGTCTCTGCTCAAGCGCGGGCT 61531
212 .....AsnThrProIleProGlnAspThrArgG1 221
61532 CGTCTCTTCTTTTTCACGGTCTTGGCGCTCGCCCGAGGGCTTCCGTT 61581
221 yLeuSerGluLysSerLeuGlnLysSerAlaLysValValTyrIleLeuG 238
61582 CTGTGCGGAGAGTCTGAGTGGAGCGATCTGGAAATCTGGCCATGGCG 61631
238 luLysLysHisSerArgAlaAlaThrGlyIleLeuLysLeuLeuAlaAsp 254
61632 AGCATCGACGCTCGCTCCGAGGACGCGCTGGTAAGAAGTACGTCCTCA 61681
255 LysAsnSerAspLeuPheLysLysTyrAlaLeuPheSerProSerAspH1 271
61682 GAGCGCGCTCGCTGATTCAGCACTATGCA..... 61711
271 sArgValProArgIleTyrValProLeuLysAspCysProGlnAspPheM 288

```

61712 .....GAGGCACCG..... 61720  
288 eThrArgProLysAspPheAlaAsnThrLeuPheIleCysArgIleIle 304  
||||| .....TTAGTA 61732  
61721 .....TTTGGC..... 315  
305 AspTrpLysGluAspCysAsnPheAlaLeuGly..... 315  
61733 GGTTCACGCTGGGTTCGGGTTCGTCATGGGGACACCGTGGAGTCCG 61782  
316 .....GlnLeuAlaLysS 320  
61783 CAGTCGTTCGGGCGACCGGCTCGCTGGCGCTCTTCACGTTGGGGCGCA 61832  
320 eLeuGlyGlnAlaGlyLeuIleGluProGluThrGluGlyIleLeuThr 336  
||||| ..... 61879  
61833 GCTTCATCTCTTCTCAGATACCCCGGAGTGGACCGCATATA... 61879  
337 GluTyrGlyValAspPheSerAspPheSerSerGluVal.....LeuG 351  
61880 .....ATAGCAAGCTCTTCTTCGAAATGCGCGGTTT 61914  
351 LuCysLeuPro..... 354  
61915 CGTGCATCCACCACCAAGTCCAGCCCATGCTCCGCCACACAAGTCA 61964  
355 GlnSerLeuProTrpThrIleProProAspGluValGlyLysArgArgAs 371  
61965 TCACAGACACCATGTCGCTCCGGCCCGCGGACTCGAAGAGACCGCC 62014  
371 pLeuArgLysAspCysIlePheThrIleAspProSerThrAlaArgAspL 388  
62015 GTGAGATCGGGTCCCTATCTGTCG.....CATCGCGGCTCGGACGA 62058  
388 euAspAlaLeuAlaCysArgArgLeuThrAspGlyThrPheGluVal 404  
62059 TG.....TGATCGTGGCTCCGAGCGACGTTCCAGGATCTACAATCTC 62099  
405 GlyValHisIleAlaAspValSerTyrPheValProGluGlySerSerLe 421  
62100 GCACACGGGCGCTCTATATCATCTCTTCCCGGAGATCAGCAATT 62149  
421 uAspLysValAlaLaGluArgAlaThrSerValTyrLeuValGlnLysV 438  
62150 CTGCTCGATCAGGCGCGAGATCATGCATC.....G 62183  
438 alValProMetLeuProArgLeuLeuCysGluGluLeuCysSerLeuAsn 454  
62184 TCGACTCGCATCTCAATCCGCTGCTCGCGGAGGACGACGTCGTACGGC 62233  
455 ProMetThrAspLysLeuThrPheSerValIleTrpLysLeuThrProGl 471  
62234 CCGCGCTTCGAG.....GCAAAATGATGGCAG.....CCTCC 62265  
471 uGlyLysIleLeuGluGluTrpPheGlyArg..... 481  
62266 CTCGGCGCGCGAGATGTTGGGAGCAGCGTGGGCGCTGGCGCGCGCG 62315  
482 .....ThrIleIleArgSerCysThrLysLeuSer 491  
62316 CAGGCGCGGAGGCGCATGAGCCCTTCTGGACGTTTGCAGT.....ATA 62359  
492 TyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysIleProGl 508  
62360 GGAGATTTTATGACACAGGACGACGCAATCAGAGTGAGACGACCTGC 62409  
508 uGluGluLeuProProIleSerProGluHis.....S 519  
62410 TTTTCGATTTCAAGCCGTTTCGCGCTGGGTACGCGGAGGACCGGTTCCCGC 62459  
519 eValGluGluValHisGlnAla..... 526  
62460 CGATCGAGCGCTGAGAGGCGAACCCCATCTTCTACTGGGATGAAGGC 62509

527 .....ValLeuAsnLeuHisSerIleAlaLysGlnLeuArgAr 539  
62510 CGTCTCGGTCTCATCCGACGAGTGTGCGGGTGTTCGCGGA 62559  
539 gGlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlap 556  
62560 CGAACGCTTCGCG.....GTCAGTCGAGAAGAGTGGGAAT 62594  
556 hethrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleIleYrclu 572  
62595 CGAGCGGAGTACTCGTGGCCATTCGCGAG..... 62626  
573 TyrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsnMe 589  
62626 ..... 62626  
589 tAlaValAlaHisLysIlePheArgThrPheProGluGlnAlaLeuLeuA 606  
62626 ..... 62626  
606 rgArgHisProProGlnThrLysMetLeuSerAspLeuValGluPhe 622  
62627 .....CTCAGCGATATGAAGAAGTAC 62647  
623 CysAspGlnMetClyLeuProMetAspValSerSerAlaGlyAlaLeuAs 639  
62648 ...CGATTGTTCGGGCTCGCGCGGAG.....GATCAGCTCGGTCCG 62688  
639 nLysSerLeuThrLysThrPheGlyAspAspLysTyrSerLeuAlaArgL 656  
62689 CAAGCTCGTCAACCCGCTTTTACGTCACGCGCATCGACCTGCTCGCGG 62738  
656 ysGluValLeuThrAsnMetTyrSerArgProMetGlnMetAlaLeuTyr 672  
62739 CGAAATA.....CAGCCACCGTCCGACGAGTGTCTGAT 62773  
673 PheCysSerGlyMetLeuGlnAspGlnGluPhe.....ArgHi 686  
62774 GCTCGCTCCGGA.....CAAGAGAGTTCGAGCTTGTGCGGA 62811  
686 sTyrAlaLeuAsnValProLeuTyr..... 694  
62812 TTACCGGAGGGAATCCCGATCGCGCGCATCAGCGCTCTGTTCAAGGTTTC 62861  
695 ..ThrHisPheThrSerProIleArgArgPheAlaAspValIleValHis 710  
62862 CGGCGGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGCG..... 62902  
711 ArgLeuLeuAlaAlaLeuGlyTyrSerGluGlnProAspValGluPr 727  
62903 .....ACTGCGCGCGCTCGCGTGGGTTTGGTCCCGCAGGTCGATGA 62946  
727 oAspThr..... 729  
62947 GGAGACCAAGACCGTGTGCTCGCTCAGCGAGGGGCTCGCGCTGCTCC 62996  
729 ..... 729  
62997 ATGAGCTCTCGATGAGCGCGCAGGAACCCGCTCGAAATGACGCTCTTG 63046  
730 .....LeuGlnLysGlnAlaAspHisCysAsnAspArgArgMetal 743  
63047 ACGATGCTGTCTTACGCGGAGCGCGAC.....GCCAGCAGGCTGAG 63087  
743 aSerLysArgValGlnGluLeuSerIleGly..... 753  
63088 CACGAGAGGAGTGTGCGCGTCTC...GTGGTGGATTATCGCTGTGGCA 63134  
754 .....LeuPhePheAlaVal.....LeuValLys 761  
63135 CGGATACCAACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTCGCG 63184

```
762 GluSerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnG1 778
    :::: ||||| |||||:
63185 TGGCCGAGGCGCTCGAGCTGGTGAAGCGCGAGCCGGCTCATGAGGAA 63234
    :|||: ||||| |||||:
778 alaPheaspValLeuVal.....L 785
    :|||: ||||| |||||:
63235 CGCGTCGATGAGTGTCTCGCTCGCATATATCCTCAGAAATAGGAATG 63284
    :|||: ||||| |||||:
785 euArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArg 801
    :|||: ||||| |||||:
63285 TGGCTTTGGCAGCGACCGCTGGAGTACTGGGGGATCGATC..... 63328
    :|||: ||||| |||||:
802 SerTyrSerPheGlnLysValGlyLysLysProGlnLeuThrLeuValTr 818
    :|||: ||||| |||||:
63329 .....AAGAAAGGGAGATGGTCTTCTCT 63354
    :|||: ||||| |||||:
818 pGluProAspLeuGluGluProThrGlnGlnValIleThrIleP 835
    :|||: ||||| |||||:
63355 GATCCGAGCGCTCGAGATGG.....ACTGTAT 63386
    :|||: ||||| |||||:
835 heSerLeuValAspVal 840
    :|||: ||||| |||||:
63387 TCTCAGCCAGACGTG 63403

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-480-1

seq_documentation_block:
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molinar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1

alignment_scores:
    Quality: 121.50      Length: 1023
    Ratio: 0.292
Percent Similarity: 40.665 Percent Identity: 18.084

alignment_block:
US-09-647-377-9 x US-09-568-480-1 ..

Align seq 1/1 to: US-09-568-480-1 from: 1 to: 68750

11 ArgSerProGlyThrProArg.....GlyValSerSera 22
    :|||: ||||| |||||:
60857 AGGGCCAGGGCTCGCGGCTCGAGTCACTGGGGCATGTTCCACGA 60906
    :|||: ||||| |||||:
22 lValGlyProSerAlaValGlyAlaSerProGlyAspLysSerIysA 39
    :|||: ||||| |||||:
60907 GGTGGGGATGGCGGTGGCAAGAAACCGTGGCGCGGCGAGATC...T 60953
    :|||: ||||| |||||:
39 snLysSerMetArgGlyLysLysSerIlePheGluThrTyrMetSer 55
    :|||: ||||| |||||:
60954 CTCGGGGATGGCGGC.....ATCAC 60976
    :|||: ||||| |||||:
```

```
56 LysGluAspValSerGluGlyLeuLysArg.....GlyThrLeuI1 69
    :|||: ||||| |||||:
60977 CCGGATGAGGGTCTGTGAGCTCTGGCGCGTCTCGAGGGTGATCGCGT 61026
    :|||: ||||| |||||:
69 eGln...GlyValLeuArgIleAsnProLysLysPheHisGluAlaPheI 85
    :|||: ||||| |||||:
61027 GCAGACGGGGTGATACCGATCACTCCGCGCGAGTGGTGGAG...TTCT 61073
    :|||: ||||| |||||:
85 leProSer..... 87
    :|||: ||||| |||||:
61074 ACCCGCAACAGCGGCTCACGAGGTTGTGCGGCTGTGACCAACGAG 61123
    :|||: ||||| |||||:
88 .....ProAspGlyAspArgAspIlePheIleAspG1 98
    :|||: ||||| |||||:
61124 CGCGCGGTGCTGATCGGACCGCGGGATCGGACCTGCTCGACAGCT 61173
    :|||: ||||| |||||:
98 yValValAlaArgAsnArgAlaLeuAsnGly.....AspLeuValY 112
    :|||: ||||| |||||:
61174 TGGGTGCGGTGAGCGAGCGCGCGGGGCTGTGCGAGGACGTCGTGC 61223
    :|||: ||||| |||||:
112 alValLysLeu.....LeuProGluAspGlnTrpLysAla 123
    :|||: ||||| |||||:
61224 GCGTGCAGGTCTGCAATGTGCTGCTCCCTGAAGACAAGATCGAGGTG 61273
    :|||: ||||| |||||:
124 valLysProGluSerAsn..... 129
    :|||: ||||| |||||:
61274 GATGCCCGGCTCTCGAGCATGGGCTGCTGCTGATGACCTGGAGCT 61323
    :|||: ||||| |||||:
130 .AspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGluYC 146
    :|||: ||||| |||||:
61324 GCGCAACCGCATCGAGCTGCTGGGCGTCCGCGGCTCGACCCCTGG 61373
    :|||: ||||| |||||:
146 ysGlyHisHisProGluGlnSerArgLysGlyTyrSerGlyProAsp 162
    :|||: ||||| |||||:
61374 GTGGACGTACCCACGATAGCAGCATACGCGCTGCTCGACGAC 61423
    :|||: ||||| |||||:
163 ValIleGluAlaGlnPheAspAspSerAspSerGluAspArgHisG1 179
    :|||: ||||| |||||:
61424 GCCCTCGTCTCGGCTTGGCGCGGTCGACACGAGCAATCGAGCGC 61473
    :|||: ||||| |||||:
179 yAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerThrP 196
    :|||: ||||| |||||:
61474 GAGCGCGGTTCTGCTGCTCCACGCTCCTCGC..... 61504
    :|||: ||||| |||||:
196 roAspArgGlyLysGluAspSerSerThrProValMetLysAspGlu... 211
    :|||: ||||| |||||:
61505 .....TTTGTCTGTCTGCTCAAGCGCGGCT 61531
    :|||: ||||| |||||:
212 .....AsnThrProIleProGlnAspThrArgG1 221
    :|||: ||||| |||||:
61532 CGTCTCTCTGTTTACGGTCTCGCGGCTCGCGGAGGCTCGCGTTC 61581
    :|||: ||||| |||||:
221 yLeuSerGluLysSerLeuGlnLysSerAlaLysValValTyrIleLeuG 238
    :|||: ||||| |||||:
61582 CTGGTCGGAGAAGTCTGAGTGGAGGATCTGGAATCGTGCCATGTGCG 61631
    :|||: ||||| |||||:
238 luLysLysHisSerArgAlaAlaThrGlyIleLeuLysLeuAlaAsp 254
    :|||: ||||| |||||:
61632 AGCATCGAGCGCTCGCTCGAGGAGCGCTGTGAAGAGTACGTCCAA 61681
    :|||: ||||| |||||:
255 LysAsnSerAspLeuPheLysLysTyrAlaLeuPheSerProSerAspHi 271
    :|||: ||||| |||||:
61682 GAGCGCGCTCGTGTATTCAGCACTATGCA..... 61711
    :|||: ||||| |||||:
271 sArgValProArgIleTyrValProLeuLysAspCysProGlnAspPheM 288
    :|||: ||||| |||||:
61712 .....GACGACCG..... 61720
    :|||: ||||| |||||:
288 etThrArgProLysAspPheAlaAsnThrLeuPheIleCysArgIleIle 304
    :|||: ||||| |||||:
61721 .....TTTGG.....TTAGTA 61732
    :|||: ||||| |||||:
305 AspTrpLysGluAspCysAsnPheAlaLeuGly..... 315
    :|||: ||||| |||||:
```



```

785 euArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArg 801
      ::|||||::: ||| |||||:::||||
63285 TGCGTTCCGAGCAGGACCTGGAGTACTCGGGGCATCGATC..... 63328

802 SerTyrSerPheGlnLysValGlyLysLysProGluLeuThrLeuValTr 818
      ||||| |||||:::|||||
63329 .....AAGAAAGGGGAGATGGTCTTCTCCT 63354

818 pGluProAspAspLeuGluGluProThrGlnGlnValIleThrIleP 835
      ||||| |||||:::|||||
63355 GATCCGCGCGCCCTGAGATGG.....ACTGTAT 63386

835 heSerLeuValAspVal 840
      ||||| |||||
63387 TCTCCAGCCAGACGTG 63403

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-486-1

seq_documentation_block:
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

```

```

alignment_scores:
  Quality: 121.50      Length: 1023
  Ratio: 0.292        Gaps: 46
  Percent Similarity: 40.665      Percent Identity: 18.084

alignment_block:
US-09-647-377-9 x US-09-568-486-1 ..

Align seg 1/1 to: US-09-568-486-1 from: 1 to: 68750

11 ArgSerProGlyThrProArg.....GlyValSerSerVa 22
   |||||::: ||||| |||||::: |||||::: |||||:::
60857 AGGCGGAGGCGCTGCGCGCTGAGATCGACTGGGGCATGTTCACGA 60906

22 IvalcylProSerAlaValcylAlaserProGlyAspLysSerLysA 39
   |||||::: |||||::: |||||::: |||||::: |||||:::
60907 GGTGGGATGGCGTTTGCAGAAAACCGTGGCGCGCGCAGATC...T 60953

39 snLysSerMetArgGlyLysLysLysSerIlePheGluThrTyrMetSer 55
   ::::::::::: ||||| |||||
60954 CTCGGGATGCGGGC.....ATCAC 60976

56 LysGluAspValSerGluGlyLeuLysArg.....GlyThrLeuI 69
   ::::::: |||||::: |||||
60977 CCGCATGAGGCTGTGTCAGCTTGGCGGCTTGTCTGAGGGGTATCGCT 61026

69 eGln...GlyValLeuArgIleAsnProLysLysPheHisGluAlaPheI 85
   ::||| |||||::: |||||::: |||||::: |||||
61027 GCAGCGGGGTGATACCGATCACTCCGCGCAGTGGGTGGAG...TTCT 61073

```

```

85 leProSer..... 87
   |||||:::
61074 ACCCGGAACAGCGGCTCACGGAGTTGTGGGGTGTGTGACACGAG 61123

88 .....ProAspGlyAspArgAspIlePheIleAspS1 98
   ||||| |||||::: |||||::: |||||:::
61124 CGCGCGTCTGTGATCGGACCCCGGGGATCGGACCTGCTCGAACAGCT 61173

98 yValValAlaArgAsnArgAlaLeuAsnGly.....AspLeuValV 112
   :::: ||||| ||||| ||||| |||||
61174 TGGCTCGCTGAGCCGAGCGCGGGGCTGCTCGAGGACGCTGCTGC 61223

112 alValLysLeu.....LeuProGluAspGlnTrpLysAla 123
   |||||::: ||||| ||||| ||||| |||||
61224 CGGTGAGGTCTGCGATGTGCTGCTCTCCCTGAGCAAGATCGAGGTG 61273

124 ValLysProGluSerAsn..... 129
   ||||| |||||
61274 GATGCCCGGCTCTCGAGCATGGCATGGCATGCTGATGAGCCTGAGCT 61323

130 .AspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGluGlyC 146
   :::::: ||||| ||||| ||||| |||||
61324 GCGCAACCGCATCGAGGCTGCTGGGGCTCGCGCGCTCGAGCCTGG 61373

146 ySGlyHisProLeuGlnGlnSerArgLysGlyTrpSerGlyProasp 162
   :::::: ||||| ||||| ||||| |||||
61374 GTGGAGCTACCCCAACGGTAGCAGCGATAACGCGCTGCTGCTGACG 61423

163 ValIleIleGluAlaGlnPheAspAspSerAspSerGluAspArgHisG1 179
   :::::: ||||| ||||| ||||| |||||
61424 GCCTGCTGCTCGGCTTGGCGGGCTCGGACAGCGAGATCGACGGC 61473

179 yAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerThrP 196
   :::::: ||||| ||||| ||||| |||||
61474 GAGCGCGGTCTGCTCCACGCTCCCGC..... 61504

196 roAspArgGlyLysGluAspSerSerThrProValMetLysAspGlu... 211
   ||||| ||||| ||||| |||||
61505 .....TTTCGCTCTCTGCTCAAGCGCGGGCT 61531

212 .....AsnThrProIleProGlnAspThrArgG1 221
   :::: ||||| ||||| ||||| |||||
61532 CGTCTCTTCTGTTTACGGTTCGGCGGCTCGCGAGGGCTCCGCTTC 61581

221 yLeuSerGluLysSerLeuGlnLysSerAlaLysValTyrIleLeuG 238
   : ||||| ||||| :::::: ||||| |||||
61582 CTGGTCGAGAAATCTGAGTGGAGATCTGGAATCTGTGGCATGTGGC 61631

238 luLysLysHisSerArgAlaAlaThrGlyIleLeuLysLeuLeuAlaAsp 254
   : :::: ||||| ||||| ||||| |||||
61632 ACATCGCAGCTCCCTCCGAGACGCGCTGTTAAGAAGTACGTCCAA 61681

255 LysAsnSerAspLeuPheLysLysTyrAlaLeuPheSerProSerAspHi 271
   :::: ||||| ||||| ||||| |||||
61682 GAGCGCGCTCGCTGATTACACATATGCA..... 61711

271 sArgValProArgIleTyrValProLeuLysAspCysProGlnAspPheM 288
   ||||| ||||| ||||| |||||
61712 .....GACGCACCG..... 61720

288 etThrArgProLysAspPheAlaAsnThrLeuPheIleCysArgIleile 304
   ||||| ||||| ||||| |||||
61721 .....TTTGG.....TAGTA 61732

305 AspTrpLysGluAspCysAsnPheAlaLeuGly..... 315
   :::::: ||||| ||||| ||||| |||||
61733 GGGTTCAGCTGGGTGTCGGTTCGATGGGGACAGCGGTGGAGTCCG 61782

316 .....GlnLeuAlaLys 320
   ||||| ||||| ||||| |||||
61783 CAGTCGTTCCGGCGCACCGGCTCCGCTCGCGCTCTTACGTGGCGGCA 61832

```



63329 .....AAGAAAGGAGATGGTCTTCTCCT 63354  
 818 pGluProAspSerLeuGluGluProThrGlnGlnValIleThrIleP 835  
 63355 GATCCGCGAGCCCTGAGAGATGGG.....ACTGTAT 63386  
 835 heserLeuValAspVal 840  
 63387 TCTCCAGGCCACAGCTG 63403

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-568-472-1

seq\_documentation\_block:

; Sequence 1, Application US/09568472

; Patent No. 6358719

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,472

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 68750

; TYPE: DNA

; ORGANISM: Sorangium cellulosum

US-09-568-472-1

alignment\_scores:

Quality: 121.50 Length: 1023

Ratio: 0.292

Percent Similarity: 40.665 Percent Identity: 18.084

alignment\_block:

US-09-647-377-9 x US-09-568-472-1 ..

Align seg 1/1 to: US-09-568-472-1 from: 1 to: 68750

11 ArgSerProGlyThrProArg.....GlyValSerSerVa 22

60857 AGGGCGAGCGCTGCGCGCTGAGCATCGACTGGGCGATGTCACGGA 60906

22 IvalGlyProSerAlaValGlyAlaSerProGlyAspLysLysSerLysA 39

60907 GTGGGGATGGCGTTGCGAGAAACCGTGGCGCGCGCAGATC...T 60953

39 snLysSerMetArgGlyLysLysLysSerIlePheGluThrTyrMetSer 55

60954 CTCGGGGATGCGGGC.....ATCACC 60976

56 LysGluAspValSerGluGlyLeuLysArg.....GlyThrLeuI 69

60977 CCCGATAGGGGTGTGTCAGCTCTGGCGGCTTGCTCGAGGGGTGATCGGT 61026

69 eGln...GlyValLeuArgIleAsnProLysLysPheHisGluAlaPheI 85

61027 GCACAGGGGGTATACCGATCACTCCGCGCAGTGGGTGAG...TTCT 61073

85 leProSer..... 87

61074 ACCCGCAACAGCGGCTCACGAGGTTGTCGGGCTGTGGACACGAG 61123

88 .....ProAspGlyAspArgAspIlePheIleAspG 98

61124 CGCGCGTCTGATCGGACCGCGGGGATCGGAGACCTGCTGACACAGCT 61173

98 yValValAlaArgAsnArgAlaLeuAsnGly.....AspLeuValV 112

61174 TGCCTCGCTGAGCGCGCGCGCGGCTGCTGCAGAGACGTCTGTC 61223

112 alValLysLeu.....LeuProGluAspGlnTrpLysAla 123

61224 CGGTGAGGTCGCGCATGTGCTGCTCTCCCTGAAGACAAGATCGAGTG 61273

124 VallysProGluSerAsn..... 129

61274 GATGCCCGCTCTCGAGCATGGCGATCGACTCGCTGATGAGCTGGAGCT 61323

130 .AspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGluGlyC 146

61324 CGCAACCGCATCGAGGCTGCGTGGCGTCCGCGCGCTGCAGCCTTGG 61373

146 ySGlyHisHisProLeuGlnSerArgLysGlyTrpSerGlyProasp 162

61374 GTGGAGCTACCCCAACGGTAGCAGCATAAACGCGCTGGCTCGCAGCAC 61423

163 ValIleIleGluAlaGlnPheAspSerAspSerGluAspArgHisG 179

61424 GCCCTGCTCGCGCTTGGCGCGGTCGACACGAGCATCGACCGC 61473

179 yAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerThrP 196

61474 GAGCGCGGTTCGTCTCCACGCTCCCGC..... 61504

196 roAspArgGlyLysGluAspSerSerThrProValMetLysAspGlu... 211

61505 .....TTTCGCTCTGCTGTCACGCGCGGCT 61531

212 .....AsnThrProIleProGlnAspThrArgG 221

61532 CGTCTCTTCTTTACGGTTCGCGGCTCGCGAGGGCTCCCGTTC 61581

221 yLeuSerGluLysSerLeuGlnLysSerAlaLysValValTyrIleLeuG 238

61582 CTGGTCGAGAAAGTCTGAGTGGAGCATCTGGAATCGTGCGCATGTGGC 61631

238 luLysLysHisSerArgAlaAlaThrGlyIleLeuLysLeuLeuAlaasp 254

61632 ACATCGCAGCTCCCTCCGAGGACGCGCTGGTAAGAAGTACGTCCAA 61681

255 LysAsnSerAspLeuPheLysLysTyrAlaLeuPheSerProSerAspHi 271

61682 GAGCGCGCTCGCTGATTACACTATGCA..... 61711

271 sArgValProArgIleTyrValProLeuLysAspCysProGlnAspPheM 288

61712 .....GACGACCG..... 61720

288 etThrArgProLysAspPheAlaAsnThrLeuPheIleCysArgIleIle 304

61721 .....TTTGG.....TTAGTA 61732

305 AspTrpLysGluAspCysAsnPheAlaLeuGly..... 315

61733 GGGTCAGCTGGGTGTCGGTTCGTGCGGACAGCGCTGGAGCTCGC 61782

316 .....GlnLeuAlaLysS 320

61783 CAGTCGTTCCGGCGCACCGGCTCCGCTGGCGCTCTTACGTTGGCGGCA 61832

320 erLeuGlyGlnAlaGlyIleGluProGluThrGluGlyIleLeuThr 336

61833 GCTTGATCTCTTCTCAGATCACTCCCGGAGATGAGACCGCATATA... 61879

337 GluTyrGlyValAspPheSerAspPhe.SerSerGluVal.....LeuG 351

61880 .....ATAGCCAAGCTCTTCTTCCGAATGCCGGGTTT 61914







```
824 luGlulProthrGlnGlnValIleThrIlePheSerLeuValAspVal 840
      ::::: |||:::||||| |||||
57760 GAGATGGG.....ACTGTATTCTCCAGCGCAGCGTG 57791

841 .....ValLeuGlnAlaGluAlaThrAlaLeuLysTyr 851
      ||| ::| ::| ::| ::| ::| ::|
57792 TTTGATGTGCGAGCGACAGCGCGGCGCTCGCGTAC 57830

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:us-08-222-617A-3

seq documentation block:
; Sequence 3, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 04-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11601 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Acetomonium chrysogenum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 388..11526
; OTHER INFORMATION: /function= "Enzyme"
; OTHER INFORMATION: /product= "ACV Synthetase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8050..8052
; OTHER INFORMATION:
; OTHER INFORMATION: /note= "NNN-CCC, AGU, AGC, UCU, UCC, UCA, or UCG;
; OTHER INFORMATION: Xaa-Ala or Ser "
; US-08-222-617A-3
```

```
alignment_scores:
  Quality: 118.50      Length: 972
  Ratio: 0.296        Gaps: 44
Percent Similarity: 41.255 Percent Identity: 19.650
```

```
alignment_block:
US-09-647-377-9 x US-08-222-617A-3
Align seg 1/1 to: US-08-222-617A-3 from: 1 to: 11601

13 ProGlyThrProArgGlyValSerSerValValGlyProSerAlaValG1 29
|||||::: ::::: |||||:::
7245 CCAGGAGTGCAGAGAGCTTGTGATGCCAGATCCATCAGGACTT.. 7292
29 yAlaserProGlyAspLysLysSerLysAsnLysSerMetArgGlyLysL 46
||| |||||::: ::::: |||:::
7293 .GCATTCCAGGAGATCAACCAAGCTTCTCATGTGCAGCAGCATCCAAGC 7341
46 yslYsserIlePheGluThrTyrMetSerLysGluAspValSerGluG1 62
||| ::::: |||||::: |||||:::
7342 CGCATCCCCCTTCTCAGGCGGTGTTCACTGGGAAACGTACCGCGCAA 7391
62 yLeuLysArgGlyThrLeuIleGlnGlyValLeuArgIleAsnProLysL 79
::: ::| |||||::: |||||:::
7392 TGTCCACGAGGAGCAGCTGCTTCAGGAGTACAAGCGCGCTCGCCT.... 7437
79 yspPheHisGluAlaPheIleProSerProAspGlyAspArgAspIlePhe 95
7438 .....CTGCCCTTCG..... 7446
96 IleAspGlyValValAlaAlaArgAsnArgAlaLeuAsnGlyAspLeuVal 112
||| ::| ||||| ||
7447 .....GCGGCCAAGTTGATCTCAACGT 7469
112 lValLysLeuLeuProGluAspGlnTrpLysAlaValLysProGluSerA 129
| ::|||:::
7470 C.....ACGTTGAAAGAGAGCGGTCA 7489
129 snAspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGlu... 144
|||::: ::||| ::|||::: |||||
7490 ATTGCTCAACGTCAACTTCACTTCACTTACCGCCTCTTCGAGGAGGAG 7539
145 .....GlyCysG1 147
7540 ACGTTTCAGGGGTTCA TGGAACCTTCCATCTCTTCGACAACATGGC 7589
147 yHisHisProLeuGlnGlnSerArgLysGlyTrpSerGlyProAspValI 164
::|||::: ::||| ::|||::: |||
7590 CCACAAAGCTGACACAGCCTCTCGAAGCTGTGCTTGAAGATGGAG 7639
164 leIleGluAlaGlnPheAspAsp.....SerAspSerGluAspArgHis 178
::|||::: ::| ::| ::|||:::
7640 TGTGTAATCCAGAGCGCGACTTAACCTTCAGCCCTCAAGCGGACAGCGGA 7689
179 GlyAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerTh 195
::|||::: |||||::: |||||:::
7690 AATTCACCTCCATGGGCTCTTCGAGGAGCATC.....GTGGCCTCGAC 7730
195 rProAspArgGlyLysGluAspSerSerThrProValMetLysAspGluA 212
|||||:::
7731 CCGGAGCGGC..... 7740
212 snThrProIleProGlnAspThrArgGlyLeuSerGluLysSerLeuGln 228
||| ::|||::: |||||::: |||||:::
7741 ..ATCGCAATTGCTGACGCGCAGCGAGTCTCTGTACTCGAACTCAAC 7788
229 LysSerAla...LysValValTyrIleLeuGluLysLysHisSerArgAl 244
::| ::| ::|||::: |||||:::
7789 GAGCGGCAACACAGCTCGTACATTTGATCATCTCTTCTGCCAGTATTGT 7838
244 aAlaThrGlyIleLeuLysLeuLeuAlaAspLysAsnSerAspLeu.... 259
::||| ::| ||||| |||||::: |||||:::
7839 ACGACACGCGCATCTCTCTTTTGACACAGAGCATCATGATGTTGTA 7888
260 .....PheLysLysTyrAlaLeuPheSerProSerAsp 270
::||| ::||| ::|||
7889 TTGCTCTCTGGCAGTTTGAAGGCGGTGCGCATATGTGCCCTTGAC 7938
```

271 HisArgValPro..... 274  
 7939 CCGACATATCGTCGACAGAGACTGAGTCTATCTTGGAGAACTAGTGC 7988  
 275 .....ArgIleTyrValPro..... 279  
 7989 CAGGAGCTCATCACCACCTAGAAAGCACACGCGGAGGGAGGAACAGTCG 8038  
 279 ..... 279  
 8039 CAATGTTCCANNNGTGTCTTGACAGCCCGAGACCTAGCTGCCTC 8088  
 280 .....LeuLysAspCysProGlnAspPheMetThrArgProLysAs 293  
 8089 ACCAGCAGTCAAGAGAAACCGACACAGTCAACGCAGAAACGTCCTCA 8138  
 293 pheAlaAsnThrLeuPheIleCysArg..... 302  
 8139 CCTCGCATGTCTCATCTCTCACTCTCGGGAACACAGGCCAAGGGGG 8188  
 303 ..IleIleAspTrpLysGluAspCysAsnPheAlaLeuGlyGlnLeuAla 318  
 8189 TTCTGGTGGACACAG.....AGCGTAGTCCAGCTGCGC 8223  
 319 LysSerLeuGlyGlnAlaGlyGluIleGlu.....ProGluThrGI 332  
 8224 AATTCCTC.....ATCGAGCGATACTTGGGCGAGACCAA 8258  
 332 uGly.....IleLeuThrGluTyrGlyValAspPheSerA 344  
 8259 CGGGTCTCGCGCGTCTCTCTCTGCTCAACTACGTCCTCGACTCTCT. 8307  
 344 spPheSerGluValLeuGluCysLeuProGluSerLeu..... 357  
 8308 .....CTTGAACAGCTCTGTCTCTCAGTCTTGGGTGGA 8340  
 358 ...ProTrpThrIleProAspGlu..... 365  
 8341 AACAAAGTCTATCTCCACAGAGAGGGTCTCAGCAGAGGCAATCTTA 8390  
 366 ....ValGlyLysArgAspLeuArgLysAspCysIlePheThrIleA 381  
 8391 CGACATCGCGCGAGGAAGCTA.....TCTATCTCAGCGGA 8431  
 381 spProSerThrAlaArgAspLeuAspAlaLeuAlaCysArgArgLeu 397  
 8432 CGGCTCGGTGCTGCAGAGATTGAG.....CTCTCCGCTCTG 8469  
 398 ThrAspGlyThrPheGluValGlyValHisIleAlaAspValSerTyrPh 414  
 8470 CCGCAT.....CTTCACATGGTCACCGCTGCGGGCGA 8501  
 414 eValProGluGlySerSerLeuAspLysValAlaAlaGluArgAla.... 429  
 8502 GGAGTCCACGCTAGTCTAGTTTGAAGATCGCTCCAGTTCGCGGGCC 8551  
 430 .....ThrSerValTyrLeuVal... 435  
 8552 AGATCAACAACCGCTATGATATCACTGAGACGCGTGTACAAATCATC 8601  
 436 .....GlnLysValValProMetLeuProArgLeuLysCys..... 447  
 8602 ACCAGTTCAGGGCGATGCCCTTTACCAAGCACTCTGCGCGGAT 8651  
 448 .....GluGluLysCysSerLeuAsnProMetThrAspLysLeuThrP 462  
 8652 CCCCGAAGTCAGTCTAGCTCTGAAGACCGGACTTCAGGFGTTCCTT 8701  
 462 heSerValIleTrpLysLeuThrProGluGlyLysIleLeuGluThrP 478  
 8702 TCAACGCTGTGCGAGCTCTACTTGGCGGTGACTGCCCTGTGCGGG 8751

479 Phe.....GlyArgThrIleIleArgSerCysThrLysLeuSe 491  
 8752 TACCTCAACGAGTCCCTGACCAACGAGGATTCATCCCAACCCCTTT 8801  
 491 rTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysIleProG 508  
 8802 CTAGACCCGAACAGCAAGTGAAGTCGCTCCCCAGAGACTCTACAAGA 8851  
 508 LuGluGluLeuProProIleSerProGluHisSerValGlu..... 521  
 8852 CTGAGATCTGGTGGCTTCGCTGGGACCCCAACCATCTCGAGTATCTCGGC 8901  
 522 .....G1 522  
 8902 CGCAAGGACGACGAGGTCAAGCTGAGGGCTTCGCAATCAGCTCTCCGA 8951  
 522 uValHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgA 539  
 8952 GGTGGGATGCGCTCTAGCCATCTCTGCTGTTAAGGAGGCTGCGTCA 9001  
 539 rgGlnArgPhe...ValAspGly.....AlaLeuArgLeuAspGlnLeu 552  
 9002 TCCCAAGTATGACGAGGATGCTCCGATTCACGAAGGTCAGCGCATC 9051  
 553 LysLeuAlaPheThrLeu..... 558  
 9052 GTCTGCTACTACAGCTCAACGCGGAACCTGTGTGCGAAGCATCGAGTAT 9101  
 559 ....AspHis...GluThrGlyLeuProGln.....GlyCysHisI 570  
 9102 CGGTGACCACTGCGCAACCTTCCCGCTACATGTCCTCCCAAGTCAGA 9151  
 570 LeTyrGluTyrArgAspSerAsnLysLeuValGluGluPheMetLeuLeu 586  
 9152 TCCACAGTTGGAGGATCTCTCCCGTGACGCTCAATGGGAAGCTCGAC 9201  
 587 AlaAsnMetAlaValAlaHisLysIlePheArgThrPheProGluGlnAl 603  
 9202 CTGAACAGGCTCTCCACACTCAAGTCTCGCAG.....CCAGAGTTTA 9245  
 603 aLeuLeuArgHisProProGlnThr.....LysMetLeuSerA 618  
 9246 CACCGCTCCAGMAATTTCGACAGAGAAACCTTGTGCGAGCTTGGGCAT 9295  
 618 spLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSerSer 634  
 9296 CTCTCTTAGGC...GTCGACCACTGCGCATTCAGCAGACCTGTTGCC 9342  
 635 AlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspLysTyr 651  
 9343 CGAGGCGCGACAGCATCTCTCTCCGACTAGTGGGTGACATCTACCG 9392  
 651 rSerLeuAlaArgLysGluValLeuThrAsnMetTyr..... 663  
 9393 CCGCTAGACGCAAGGTCAACCGTCAAGGACATCTACCTCCACCGCAGCG 9442  
 663 ..... 663  
 9443 TCCGAGCCCTAAGGAAAATGCTCTGACCGACAGAGGATAAGGGTACT 9492  
 664 .....SerArgProMetGlnMetAla.....LeuTyrPh 673  
 9493 CTGCCAGGCTCTCTCCCTCCAGCGGCGGAGCAGGCCAGGTTGAGGG 9542  
 673 eCysSerGlyMetLeuGlnAspGlnGluGlnPheArgHisTyrAlaLeuA 690  
 9543 CGACGACCGCTTCTCCCATCCAGGACTGGTCTCTTCCCAAGCCCTGG 9592  
 690 snValProLeuTyrThrHisPheThrSerProIleArgPheAlaAsp 706  
 9593 ATAAACCCGCTTACTGGAACCACTGCTTCAACCATTCGAACCGGGCA... 9639  
 707 valIleValHisArgLeuLeuAlaAlaAlaLeuGlyTyrSerGluGlnPr 723



7741 ...ATFCGCAATTCGTGACGGCACCGAGGAGTCTCTCTACTCCGAACTCAAC 7788

229 LysSerAla...LysValValTyrIleLeuGluLysLysHisSerArgAl 244  
::: ||| :::::||||:::||||::: ||| :::  
7789 GAGCGGGCAACACAGCTCGTACATTTTGATCATCTCTCTGCGCATTTGT 7838

244 alaThrGlyIleLeuLysLysLeuAlaLysLysSerAspLeu... 259  
::: ||| ::: ||||| ||||| ||||| :::  
7839 AGCAGACGACCGCATCGCTCTCTTTTGGCAGAGAGCATCATGATGTGA 7888

260 .....PhelLysLysTyrAlaLeuPheSerProSer 270  
::: ||| ::: ||| ::: ||| :::  
7889 TTGCTCTCTGGCAGTTTGGGAAGCGGTCGCGCATATGTGCCCTTGAC 7938

271 HisArgValPro... 274  
||| ||| ||| |||

7939 CCGCATATCCGTCGCGAGACGTGAGCTATCTTTGGAGGAATCTAGTGC 7988

275 .....ArgIleTyrValPro... 279  
||| :::::||||  
7989 CAGGACGCTCATCACCACCTAGAAAGCACACGCCGAGGGGAGAACAGTCG 8038

279 ..... 279  
::: |||||

8039 CAATGTTCCANNNGTGTCTTGACAGCCCCGAGACCTAGCTGCCTC 8088

280 .....LeuLysAspCysProGlnAspPheMetThrArgProLysAs 293  
|||::: ||| :::::||||  
8089 AACCAAGCAGTCAAAAGGAAACCCGCAACACGTCAACGCAGAACCGTCCGA 8138

293 pPheAlaAsnThrLeuPheIleCysArg... 302  
||: ||| ||| ||| ||| ||| |||

8139 CCTCGCATATGCTATCTTCACTCGGGAACACAGCAAGCCCAAGGGGG 8188

303 ...IleIleAspTrpLysGluAspCysAsnPheAlaLeuGlyGlnLeuAla 318  
::: ||| ::::: ||| ::::: |||

8189 TTCTGGTGGAGCACAG... 8223  
::: ||| ::::: ||| ::::: |||

319 LysSerLeuGlyGlnAlaGlyGluIleLeu...ProGluThrG 332  
::: ||| ||| ||| ||| ||| ||| ||| |||

8224 RNTTCCCTC...ATCGCGGATCTTCGGCGAGCAAC 8258

332 uGly...IleLeuThrGluTyrGlyValaspPheSerA 344  
::: ||| ::::: ||| ::::: |||

8259 CGGCTCTACGCGCTCTCTCTCTGCTCCAACTACGTCCTTCGACTCTCT. 8307

344 spPheSerSerGluValLeuLysLeuProGlnSerLeu... 357  
||| ||| ||| ||| ||| ||| ||| |||

8308 .....CTTGAAGCTCTCTCTCTCAGTCTGGGTGA 8340

358 ...ProTrpThrIleProAspGlu... 365  
||| ||| ||| ||| ||| ||| ||| |||

8341 AACAGCTCATCATCTCCACGAGGAGGGTCTCAGCAGAGGCATTCCTA 8390

366 ...ValGlyLysArgAspLeuArgLysAspCysIlePheThrIleA 381  
::: ||| ||| ||| ||| ||| ||| |||

8391 CGACATCGCGCAGGAGGAAGCTA...TCTATCTCAGCGGA 8431

381 spProSerThrAlaArgAspLeuAspAspAlaLeuAlaCysArgArgLeu 397  
||| ||| ||| ||| ||| ||| ||| |||

8432 CGCCCTCGGTGCTGCAGCAGATTGAG...CTCTCCCGTCTG 8469  
||| ||| ||| ||| ||| ||| ||| |||

398 ThrAspGlyThrPheGluValGlyValHisIleAlaAspValTyrPh 414  
::: ||| ||| ||| ||| ||| ||| |||

8470 CCGCAT...CTTCACATGTGTACCGCTCGCGGCGA 8501

414 eValProGluGlySerSerLeuAspLysValAlaAlaGluArgAla... 429  
::: ||| ||| ||| ||| ||| ||| ||| |||

8502 GGAGTTCACGCTAGTCAGTTTGAGAAGATCGCTCCGACTCGCGGGCC 8551

430 .....ThrSerValTyrLeuVal... 435  
||| ||| ||| ||| ||| ||| ||| |||

8552 AGATCAACAACCGCTATGTTATCATCTGAGACGACCGGTGACACATCATC 8601

[illegible]

```

664      ....SerArgProMetGlnMetala.....LeutyRph 673
9493  CTGCCAGCGTCTCTCCCTCCAGCGAGCGGAGCCAGGTTGAGG 9542
673  eCysSerGlyMetLeuGlnAspGlnGluPheArgHisTyraLeuA 690
9543  CGAGCACCGCTTCTCCCATCCAGGACTGGTTCTTCCAGAGCCCTGG 9592
690  snValProLeuTyThrHisPheThrSerProIleArgArgPheAlaA 706
9593  ATAAACCCCGCTTACTGGAACCACTGCTTCAACCATCGAACCGGGCA... 9639
707  ValIleValHisArgLeuLeuAlaAlaLeuGlyTySerGluGlnPr 723
9640  CTCCTCCGTCGAAGGGCTCGGGGTGCTCTGAAGCTGTGTCAGAGCGCCA 9689
723  oAspValClnProAspThrLeuGlnLysGln..... 733
9690  CGAGCTGCTGGTCTGAGACTGCAACGGCGGAGCAAGGTGCCATGTTTC 9739
734  ....AlaAspHisCysAsnAspArgAlaMetala..... 743
9740  AGACCTTTCGGCTGACTGCGCGCAACCTCGCTTGACTGTCTAGACCGA 9789
744  ....SerLysArgValGlnGlu.....LeuSerIleGl 753
9790  CGAAGCTTCGAGGACCGAGGATGTACAGGAGCTCTCTCCGAGATCCA 9839
753  yLeuPhePheAlaValLeuValLysGluSerGlyProLeuGluSerGluA 770
9840  ATCTCATTTGCAGCTC.....GAGAATGACCCCTCTACACAGCTGG 9880
770  laMetValMetGly 774
9881  CGTACATCCAGGT 9894

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-845-258-9

seq_documentation_block:
; Sequence 9, Application US/08845258
; Patent No. 6183976
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; OF INFECTION: AND TREATMENT OF B. MICROPI INFECTION
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPED AND BERRY
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,258
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 9

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 07:57:53 ; Search time 2321.88 seconds  
(without alignments)  
18020.122 Million cell updates/sec

Title: US-09-647-377-8  
Perfect score: 3100  
Sequence: 1 cggcgccgcgcctcccg.....tgaaaaaaaaaaaaaaaaaaaa 3100

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_gss:\*
  - 13: em\_gss\_hum:\*
  - 14: em\_gss\_inv:\*
  - 15: em\_gss\_pin:\*
  - 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	674.4	21.8	695	10	BI078196
2	672.6	21.7	701	10	BG864079
3	671	21.6	681	10	BF450541
4	668.2	21.6	1158	10	BI650394
5	647	20.9	682	10	BI661242
6	643.8	20.8	674	9	BB037094
7	620.2	20.0	898	10	BF182212
8	619.6	20.0	725	10	BE380437
9	590.2	19.0	887	10	BE791331
10	586.4	18.9	600	10	BI988256
11	583.4	18.8	600	10	BG804017
12	579	18.7	622	10	BE284271
13	570.8	18.4	735	10	BE825206
14	566	18.3	1127	10	BG746662
15	538.2	17.4	1161	10	BG106547
16	526.2	17.0	961	10	BM012844
17	518.8	16.7	666	9	BEL11729

18	518.8	16.7	711	10	BG750784
19	509	16.4	521	10	BE534295
20	499.4	16.1	643	9	BB620465
21	498.4	16.1	692	9	BB615143
22	491.8	15.9	825	10	B1114256
23	491.4	15.9	684	9	BB636311
24	486.4	15.7	546	9	BB751672
25	481.2	15.5	697	10	BE545087
26	480	15.5	529	9	BB771440
27	477	15.4	761	10	BG572376
28	471.8	15.2	494	9	AA756851
29	471.4	15.2	506	10	BE910893
30	457.6	14.8	573	9	BE110335
31	449	14.5	451	9	AI597514
32	443.2	14.3	887	10	BF304859
33	443	14.3	444	10	BE380350
34	440.2	14.2	560	9	BE096022
35	431.8	13.9	439	9	BB821152
36	431.8	13.9	468	9	BB822910
37	426.8	13.8	483	10	BF548574
38	425.6	13.7	588	10	BF198482
39	424.8	13.7	439	9	AA575234
40	416.2	13.4	427	9	AI464025
41	416	13.4	425	9	AI585294
42	412.8	13.3	456	9	AA553101
43	412.2	13.3	476	9	BB761195
44	412	13.3	417	9	BB751180
45	404.6	13.1	693	10	BF204262

ALIGNMENTS

RESULT 1

BI078196  
LOCUS 602872520F1 NCL\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5004341 5', linear EST 20-JUN-2001  
DEFINITION mRNA sequence.  
ACCESSION BI078196  
VERSION BI078196.1 GI:14496526  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: L1AM11042 row: h column: 06  
High quality sequence stop: 689.  
Location/Qualifiers  
1. .695  
/organism="Mus musculus"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5004341"  
/clone\_lib="NCL\_CGAP Mam2"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="5 months"  
/lab\_host="PH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: Sali; Site\_2: Noti; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator

FEATURES  
source



```
QY 1645 tgaggagagcttcccccaatttccagagcacagcgtcgaggaggtgcaccagggcagt 1704
|||||
Db 541 TGAGAAAGAGCTTCCCCCAATTTCTCCAGAGCACAGCGTCGAGGAGGTGCACAGGGCAGT 600
QY 1705 cctgaacctcacagcattgcaaa--gcaactccgcgcagcgcgtttgttagatggcgca 1762
|||||
Db 601 CCTGAACCTGCACAGAGTTGCAAGGCCAACTCCGAGCGCCAGCGCTCTGTCAATGGCGCA 660
QY 1763 ctccgttttagatcagctgaagcttgcttttactcttgagacca 1803
|||||
Db 661 CTCGGTTTAGATCAGTTGAAGCTTGCTTTATCTGTGAGCA 701

RESULT 3
LOCUS BF450541 681 bp mRNA linear EST 29-DEC-2000
DEFINITION uz67f12.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3674159 5'
similar to SW:YR86_CAEEL 009368 HYPOTHETICAL 93.7 KDA PROTEIN
F48E8.6 IN CHROMOSOME III. [2] SW:YR86_CAEEL ; mRNA sequence.
ACCESSION BF450541
VERSION BF450541.1 GI:11516710
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 681)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
MGI:1434927
Seq primer: -40RP from Gibco
High quality sequence stop: 440.
Location/Qualifiers
1. .681
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3674159"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 160 a 206 c 169 g 144 t 2 others
ORIGIN

Query Match 21.6%; Score 671; DB 10; Length 681;
Best Local Similarity 99.0%; Pred. No. 2.6e-106;
Matches 674; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1589 aaactgagctacgacctgccagagcatgatcgaaatccaaactgagaagatccctgag 1648
|||||
Db 1 AAACCTGAGCTACGACCATGCGCCAGAGCATGATCGAAATCCAACTGAGAAGATCCCTGAG 60
QY 1649 gaagagcttcccccaatttccagagcacagcgtcgaggaggtgcaccagggcagtcctg 1708
|||||
```

```
Db 61 GAAGAGCTTCCCCCAATTTCTCCAGAGCACAGCGTCGAGAGGTGCACAGGCGAGTCCTCG 120
QY 1709 aaactgacagcattgcacaagcaactccgcgcagcgcctttgttagatggcgcaactccgt 1768
|||||
Db 121 AACCTGCACAGCATGTGAAAGCAACTCCGCCCGCAGCGCTTTGTAGATGGCGCAGCTCCGT 180
QY 1769 ttagatccagctgaagcttgcttttactctggaccatgagactggactggactgctcaagatgt 1828
|||||
Db 181 TTAGATCAGCTGAAGCTTGCTTTTACTCTGGACCATGAGACTGGATTGCCCTCAAGGATGT 240
QY 1829 cacatctatgagtaccagacagcaacaagcctgttagagaggttcattgctcctcgccaac 1888
|||||
Db 241 CACATCTATGAGTACCGGAGACAGCAAAAGCTGTGTAGAGGAGTTTCATGCTCCTGSGCAAC 300
QY 1889 atggcggtggcccaacaagactcttcgcacacttccctgagcaggccctgctgcgcgcgcgt 1948
|||||
Db 301 ATGGCGGTGGCCCAAGATCTTCCGCACCTTCCCTGAGCAGGCCCTGCTCGCCCGCAT 360
QY 1949 ccccccacacagacagaagatgctcagtgacctgtgtgaggttctctatgacacagatgggctg 2008
|||||
Db 361 CCCCCACACAGACGAAGATGCTCAGTGACCTGGTGAGTCTGTGTGACCAAGATGGGCGCTG 420
QY 2009 cccatgagtctcagctctgcaggggccctaaataaaagcctgactaagacatttggagat 2068
|||||
Db 421 CCCATGGATGTGAGCTCTGCGAGGGGCCCTAAATAAAAGCCTGACTTAAGACATTTGGAGAT 480
QY 2069 gacaaatactcttgcccggaagaggtgctcaacacatgactcctccgcgcctgagcag 2128
|||||
Db 481 GACAAGTACTCTCTGGCCCGGAAGAGGTGCTCACCACATGTACTCCCGGCCCATGCGAG 540
QY 2129 atggcactgtactctgctggatgctgcaggaccagagcaggttccggcattatgct 2188
|||||
Db 541 ATGGCACTGTACTTCTGCTCTGGGATGCTGCAGGACCAAGAGACAGTTNCGGCATTATGCT 600
QY 2189 ctcaagttccctctacacacttcaacttcaacttccatccatccgccttctgtgacgcatca 2248
|||||
Db 601 CTCGAAGTTCCTCTACACACATTCACCTTCCATCCCCGCGCTTTGCTGACGICATA 660
QY 2249 gtgcacgcctcctggctgct 2269
|||||
Db 661 GTGCACCGCTNCTGGCTGCT 681

RESULT 4
LOCUS BI650394 1158 bp mRNA linear EST 12-SEP-2001
DEFINITION 603296125F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5336719 5',
mRNA sequence.
ACCESSION BI650394
VERSION BI650394.1 GI:15564630
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1158)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1855 row: a column: 08
High quality sequence stop: 762.
Location/Qualifiers
1. .1158
/organism="Mus musculus"
```

```
/db_xref="taxon:10090"  
/clone="IMAGE:5336719"  
/clone_lib="NIH_CGAP_Mam3"  
/tissue_type="tumor, gross tissue"  
/lab_host="DH10B"  
/note="organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;  
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert 2 kb. Library constructed by Life  
Technologies, catalog #12017-018. Investigators providing  
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference  
for transgenic model: Xu et al., Nature Genetics 22, 37-43  
(1999). Note: this is a NCI_CGAP Library."  
BASE COUNT      351 a   259 c   326 g   221 t      1 others  
ORIGIN
```

```
Query Match      21.6%; Score 668.2; DB 10; Length 1158;  
Best Local Similarity 94.8%; Pred. No. 6.4e-106;  
Matches 746; Conservative 0; Mismatches 33; Indels 8; Gaps 5;  
  
QY 1 cggcgccgcgcctccggg-gagcgagctgctgacaactgagctgctgaaggcaggag 59  
Db 6 CGCGCGCGCGCGCTCCCGGTCGAGCGAGCGCTGTGACAACTGAGCTGTGAGGCGAGG 65  
  
QY 60 gaactctgagctgaatagtagtgggtccctgaaactggagagaagaccaccttgaac 119  
Db 66 GAACCTTGAGCTGAATAGTAGTGGTCCCTGMAATCTGGAGAGAGACGCCACCTTGAAC 125  
  
QY 120 cagtaatgaacatactgactacaagctgaacctcgctctccggggccccccagagtg 179  
Db 126 CAGTAATGAACCATCTGACTACAAAGTGAACCTTCGGTCTCCGGGGAGCCGCCAGAGTG 185  
  
QY 180 tgtctctgtggtggccgcgagtgctgtgtgtcttcgcgcaggtgacaaaaagtcaaga 239  
Db 186 TGTCTCTGTGTGGTGGCCGAGTGTCTGTGTGCTTCGCCAGTGCACAAAAGTCAAGA 245  
  
QY 240 acaagtcctatcgaggagaagaaaagagacatattgaacctacatgtccaaaggagtg 299  
Db 246 ACAAGTCCATGCGAGGAAGAAAAGAGCATATTTTGAACCTTACATGTCTCAAGAGGATG 305  
  
QY 300 ttccagaagcttgaggagagaacacttatccagggtgtattgagaatcaacccaaga 359  
Db 306 TTTCAGAAGCTTGAAGAGAGAACACTTATCCAGGCTATTGAGATCAACCCCAAGA 365  
  
QY 360 agttcatgaagccttactctccgagatgtatggagacatattttattgatggag 419  
Db 366 ACTTTCATGAGCCCTTCAITCCTTCCGATGGTATCGGACATTTTATTATGATGGAG 425  
  
QY 420 ttgttctgtaatagagccttaaatgggaccttgggttgtaaaactgcttctcgagg 479  
Db 426 TTGTTCTCGTAATAGAGCCTTAATGGGACCTTGTGTTGTAAAACTGCTTCTGTAGG 485  
  
QY 480 atcagtggaagcagttaaaccagagagaatgacaaagaataagaagctacttatgaag 539  
Db 486 ATCAGTGAAGGCAGCTTAACACAGAGAGCAATGACAAAGAATAGAGCTACTATTAGAAG 545  
  
QY 540 ctgacatccctgaagagggtgtgacataccccctgcagcagtcctcgaaaggctgg- 598  
Db 546 CTGACATCCCTGAAGAGGGCTGTGGACATCACCCCTGCAGCAGTCCCGGAAAGGCTGGG 605  
  
QY 599 -agtggtcctgatgtcattatagagcctcagtttgatgacagcactcagaagaatagaca 657  
Db 606 AGTTGTCTGTATGTCATTATAGAGGCTTCAGTTGATACAGCCGACCTCAGAAGATAGACA 665  
  
QY 658 tggcaacaccagtggcctggttgatggtgttaaga--aatgtcaattctct-actcctga 714  
Db 666 TGSCAACACCAAGTGCGCTGGTTGATGTGTATTGCAACACTTGTCAACTCTCTAATCCTGA 725  
  
QY 715 cagagg--aaaagaagattctagtactccagttatgaagatgagaaccccccatacca 772  
Db 726 CAGAGGCAAAACACAGATCTTAGTACTCCAGTTATGAACAGATGAGAACCCACCCCATTA 785  
  
QY 773 caggaca 779
```

```
Db 786 TACAACA 792  
| |||  
RESULT      5  
BI661242  
LOCUS      603304587F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350461 5',  
DEFINITION mRNA sequence.  
ACCESSION  BI661242  
VERSION     BI661242.1 GI:15575478  
KEYWORDS    EST.  
SOURCE      house mouse.  
ORGANISM    Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  1 (bases 1 to 682)  
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.  
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL   Unpublished (1999)  
COMMENT   Contact: Robert Strausberg, Ph.D.  
          Email: cgabbs-r@mail.nih.gov  
          Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
          Ph.D.  
          cDNA Library Preparation: Life Technologies, Inc.  
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
          Clone distribution: Incyte Genomics, Inc.  
          found through the I.M.A.G.E. Consortium/LLNL at:  
          http://image.llnl.gov  
          plate: LLAM11890 row: m column: 22  
          High quality sequence stop: 682.  
          Location/Qualifiers  
            1..682  
              /organism="Mus musculus"  
              /db_xref="taxon:10090"  
              /clone="IMAGE:5350461"  
              /clone_lib="NIH_CGAP_Mam4"  
              /tissue_type="tumor, gross tissue"  
              /lab_host="DH10B"  
              /note="organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;  
              Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
              Average insert 2.5 kb. Library constructed by Life  
              Technologies, catalog # 12018-016. Investigators providing  
              samples: Lothar Hennighausen/Priscilla Furth, NIH  
              Reference for transgenic model: Li et al., Cell Growth and  
              Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP  
              Library."
```

```
BASE COUNT      195 a   140 c   187 g   159 t      1 others  
ORIGIN
```

```
Query Match      20.9%; Score 647; DB 10; Length 682;  
Best Local Similarity 99.7%; Pred. No. 3.6e-102;  
Matches 669; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
  
QY 33 tgacaactgagctgagctggaaggcagaggaactctgagctgaatagtagtggtccctgaa 92  
Db 1 TGACAACCTGAGCTGCTGAGGCGAGGAGAACTCTGACCTGAATAGTAGTGGTCCCTGAA 60  
  
QY 93 tctggagagaagacgcccacttggaaaccagtaataatgaacctoct-gactacaagctgaac 151  
Db 61 TCTGGAGAAGAAGACGCCACTTTGGAACCAAGTAATGAACCACTCTGACTACAAAGCTGAAC 120  
  
QY 152 cttcggtctccgggggaccgccagaggtgttcctctgtggttggccagtgctgttgggt 211  
Db 121 CTTCGGTCTCCGGGACCCCGAGAGGTGTCTCTCTGTGGTGGCCGAGTGTGTTGGT 180  
  
QY 212 gcttcgccaagtgacaaaaagtcaaaagtcacaagtcacatgcgaggggaagaaaagacata 271  
Db 181 GCTTCGCCAGGTGACAAAAAGTCAAAAGATCCATGCGAGGGAAGAAAAAGAGCATATA 240  
  
QY 272 ttgaaacctacatgtcccaaggaggatgttccagaaggcttgaagagaggaacacttacc 331
```

|||||  
Db 241 TTTGAACCTACATGTCCTCAAGGAGGATGTTTCAGAGGCTTGAGAGGAAACACTTATC 300  
QY 332 cagggtgtattgagaatcaaccagaagtttcattgaagccttcattctctccggt 391  
Db 301 CAGGGTGTATTGAGATCAACCAAGAGTTTCATCAAGCCCTTCATTCTCTCCGAT 360  
QY 392 ggtgatcgagacattttttattgatggagtgttctcgttaataagacccttaattgggac 451  
Db 361 GGTGATCGGACACATTTTATTGATGGAGTGTGCTGCTAATAGAGCCCTTAATGGGAC 420  
QY 452 ctgtgggtgttaaaactgctctgaggtcagtggaagcagtttaaacccagagagcaat 511  
Db 421 CTGTGTTGTAAACTGCTCTGAGGATCAGTGGAGGACGCTAAACACAGAGACAT 480  
QY 512 gacaaagaataagaacttacttatgaagctgacatccctcaagagcctgtgagacatcac 571  
Db 481 GACAAAGAAATAGAAAGTACTTATGAAGCTGACATCCCTGAAGAGGCTGTGGACATCAC 540  
QY 572 cccctgcagcagtcgccgaaagctggagtggtctcgtgatgtcattatagagctcagttt 631  
Db 541 CCCCTGCAGCAGTCCCGAAAGGCTGGAGTGGTCTGATGTCATTATAGAGGCTCAGTTT 600  
QY 632 gatgacacgactcaagaatagacatgcaacaccag-tggcctggttgatggtgttaa 690  
Db 601 GATGACGCGACTCAGAAGATAGACATGCGCAACACCAGTGGCTGGTGTGATGGTGTAA 660  
QY 691 gaattgtcaa 701  
Db 661 GAAATGTCAA 671

RESULT 6  
BB037094  
LOCUS BB037094 RIKEN full-length enriched, 13 days embryo forelimb Mus  
DEFINITION musculus cDNA clone 5930423H14 3', mRNA sequence.  
ACCESSION BB037094  
VERSION BB037094.2 GI:15403721  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 674)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y., et al. 2001  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
On Jun 10, 2000 this sequence version replaced gi:8443480.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaki,S., Inoue,K., Togawa,F., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,  
Ishii,Y. and Hayashizaki,Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.  
Funct. Genomics 2 pre, L72-L86 (2001)  
please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Location/Qualifiers  
1. :674  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="5930423H14"  
/clone\_lib="RIKEN full-length enriched, 13 days embryo  
forelimb"  
/sex="mixed"  
/tissue\_type="forelimb"  
/dev\_stage="13 days embryo"  
/lab\_host="DHL0B"

FEATURES  
source  
BASE COUNT 149 a 166 c 188 g 171 t  
ORIGIN  
Query Match 20.8%; Score 643.8; DB 9; Length 674;  
Best Local Similarity 97.5%; Pred. No. 1.3e-101;  
Matches 654; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 2414 ggccccctggagtcgagccagtcggtgagtgctcctgaaccaagcttcgacgtctg 2473  
Db 1 GCCCCCCCTGGAGTCCGAACCCAGAGGTGATGGGTTCTCTGAACCAAGCTTCAAGGTGCTG 60  
QY 2474 gtcgtgcgcttgggggtgcagaagcgcatctactcactcagtcagcctgcgacctac 2533  
Db 61 GTGTGTCGCTTTGGGGTGCAGAAAGCGCATTTAATGCAATGACGGCCCTTGGATTCCTCC 120  
QY 2534 agcttccagaaggtggggaagaagccagagctcactctgtttggagagcctgatgacctt 2593  
Db 121 AGCTTCCAGAAGTGGGGAAAAAGCCAGCAGCTCACTCTGTTTGGAGGCTGATGACCTT 180  
QY 2594 gaagaggaacacacagaggtcattcaccattctcagctggtggtggtggtggtggtgagc 2653  
Db 181 GAAGAGGAGGCCAACACAGCAGGTCATCCACCTCTCAAGCTGGGTGGATGTGGTCTGCAG 240

RIKEN full-length enriched, 13 days embryo  
primed with a primer [5'  
GAGAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 100.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGAGATCTCGATTATTAATTAATTCCTCCCTCCCTCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified plasmid vector KS(+) after bulk excision  
from Lambda FIC I. Cloning sites, 5' end: SalI; 3' end:  
BamHI."



ORGANISM	Mus musculus									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.									
	1 (bases 1 to 725)									
	NIH-MGC http://mgi.nci.nih.gov/.									
	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)									
REFERENCE	Contact: Robert Strausberg, Ph.D.									
	Email: cgaabs@mail.nih.gov									
	Tissue Procurement: Gilbert Smith, Ph.D.									
	cDNA Library Preparation: Life Technologies, Inc.									
	DNA Sequencing by: Incyte Genomics, Inc.									
TITLE	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov									
	Plate: L14M8782 row: d column: 03									
	High quality sequence stop: 596.									
	Location/Qualifiers									
	1. 725									
FEATURES	source									
	/organism="Mus musculus"									
	/strain="FVB/N"									
	/db_xref="taxon:10090"									
	/clone_lib="NCI CGAP Mam1"									
BASE COUNT	Email: cgaabs@mail.nih.gov									
	/tissue_type="tumor, biopsy sample"									
	/dev_stage="10 months, virgin"									
	/lab_host="DH10B"									
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by life Technologies. Investigator providing samples: Gilbert Smith, NIH"									
ORIGIN	206 a 179 g 167 t 1 others									
	Query Match 20.08; Score 619.6; DB 10; Length 725;									
	Best Local Similarity 97.94; Pred. No. 1.9e-97;									
	Matches 659; Conservative 0; Mismatches 10; Indels 4; Gaps 3;									
	560 tctggacatcccccctgacagctcccgaaaggctgagtggtctctgattgatttata 619									
Db	1 TGTGGACATCACCCCTCGACAGCTCCCGAAAGCTGGAGTGGCTGTGATGTCATTATA 60									
Qy	620 gaggctcattgtgacagcgactcagaagatagacatggcaacacagctggctggtt 679									
Db	61 GAGGCTCAGTTTGTATGACAGCGACTCAGAAGATAGACATGGCAACACAGTGGCTGGTT 120									
Qy	680 gatggtgttaagaaattgtcaatctctactcctcgacagaggaaagaagattctagtact 739									
Db	121 GATGGTGTAAAGAAATGTCTACTCTCTCTGACAGAGGAAAGAAAGATTCAGTACT 180									
Qy	740 ccagttatgaagatgaagaacacccccatcacacagagacagaaggctttacagaag 799									
Db	181 CCAGTTATGAAGATGAAGAACACACCCCATACACAGGACACAGAGGCTTATCAGAGA 240									
Qy	800 tcacttcagaatacagcaaaagtggtttacatctctggagaaaaagcattctcgagcagca 859									
Db	241 TCACCTCAGAAATCAGAAAGTGTGTTTACATCTTGGAGAAAAAGCATTCGACGACGA 300									
Qy	860 actggaactcctgaactcttggtgataagaacagtgactgttttaagaataacgcctg 919									
Db	301 ACTGGCATCTGAACATCTTGGCTGTATAAGAACAGTGCCTGTGTTAAGAAATACGCCCTG 360									
Qy	920 tttctccttcagaccacccagtagtaccttagtacctctctcaaggactgtccocag 979									
Db	361 TTTTCTCTTCAGACACCCAGTACCTAGAAATTTACGTACTCTCAAGGACTGTCCCCAG 420									
Qy	980 gacttcatgcccagactaaagactttgccaacgcgtgttccatctgcgcgatcatagac 1039									
Db	421 GACTTCATGACCCGACTTAAGACTTTGCCAAACACGCTGTTCATCTGCCGCATCATGAC 480									
Qy	1040 tggaaaggagactgtaattttgccctggggcgaactgggactggtcttgggcaaggctggt 1099									

Db	481	TGGAAGGAGGACTGTAATTTTCCCTGGGCAACTGGCTAAAGAGTCITGGGCAGCTGGT	540
Qy	1100	gaactcgagcctgaacagaaggatgactacagagaatagtggtgagctctctgatttc	1159
Db	541	GAATCGAGCCTGAACAGAGGATACACAGAAATATGGTGTGACTTCTCTGATCTC	600
Qy	1160	tcttcagaagttcttgaatgtctcctcaagcctccctggacaataccactgatgag	1219
Db	601	TCTTCAGAAG-TCTTGAATGTCT-CCTCANAGCCTCCGGGACAAATCC--ACTGATGAG	656
Qy	1220	gtgggcaagagaa 1232	
Db	657	GTGGGCAGGAA 669	
RESULT 9			
BE791331	601582710F1	NIH_MGC_7	887 bp mRNA linear EST 20-SEP-2000
LOCUS	mRNA sequence.		
DEFINITION	BE791331		
ACCESSION	BE791331.1		
VERSION	GI:10212529		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 887)		
	NIH-MGC http://mgi.nci.nih.gov/.		
	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
FEATURES	Email: cgaabs@mail.nih.gov		
	Tissue Procurement: DCTD/BTP		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov		
BASE COUNT	Plate: LLCM/82 row: c column: 04		
	High quality sequence stop: 750.		
	Location/Qualifiers		
	1. 887		
	/organism="Homo sapiens"		
ORIGIN	/db_xref="taxon:9606"		
	/clone="IMAGE:3937083"		
	/clone_lib="NIH_MGC_7"		
	/tissue_type="small cell carcinoma"		
	/cell_line="MGC3"		
BASE COUNT	/lab_host="DH10B (phage-resistant)"		
	/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
	238 a 205 c 229 g 215 t		
	Query Match 19.0%; Score 590.2; DB 10; Length 887;		
	Best Local Similarity 82.5%; Pred. No. 2.2e-92;		
	Matches 712; Conservative 0; Mismatches 148; Indels 3; Gaps 3;		
Qy	704	tctactcctgacagagaaagaagattcttagtactcactgataagaagtgaacaacc	763
Db	2	TGTGTTTCTGAGAAAGAGAGAGGATGTGTATGTCACCGGTTACAAAAGATGAGACACC	61
Qy	764	cccatcacacagacacaaaggccttatcagagaagtcacttcagaatacagcaagggtg	823

Db 62 TGCAATTTCAAGACACAGAGCTTTATCGGAGAAATCCCTGCAAAAGATCATCAAAAGGTG 121

Qy 824 gtttaccattggagaaaaagcattctcgagcagcaactggcatcctgaaactcttgct 883

Db 122 GTTTACATCTTGGAGAAAACATCTCGAGCAGCAACCGCTTCCTCAAACTCTGGCT 181

Qy 884 gataagaacagtgcctgtttaagaataacgcctgtttctctcttcaccaccagagta 943

Db 182 GATAAGAACAGCAACTGTTTAGGAAATAGCCCTGTGTTTCCCTCAGACACCGAGTG 241

Qy 944 cttagaatttcgtacctcgaagactgtcccgagactctcatgaccgcacctaaagc 1003

Db 242 CTTAGAAATTTATGTCCTCTCAAGAGACTGCTCCACAGACTTGTGGCAGCGCTTAAAGAT 301

Qy 1004 ttgtccaaacgcgtgttcattccgcacatcatagactggagagagactgtaattttgcc 1063

Db 302 TATGCCAACACACACTGTTTCATCTCCGCATTTGGACTGGAGAGGAGACTGCAATTTGCC 361

Qy 1064 ctggggcaactggctaaagactgttggcaggtgtgaaatcgagcctgaaacagaaagg 1123

Db 362 CTGGGCGAGCTGGCTAAAGTCTTGGCAGGCTGTGAAATTTGAGCTGAAACAGAGGA 421

Qy 1124 atactgacagaatagtgtggaactctctctctctctctcagaagttctgaatgtctc 1183

Db 422 ATACTAACAGAGTATGGCGTGATTTCTCTGATTTCTTCAGAGATTCTAGAAATGCTT 481

Qy 1184 cctcaaaagcctggcctggacaatcccaactgatgaggtggggaagagagacctaagg 1243

Db 482 CCTCAAGCCCTGCCATGGACAATTCACAGAGAGGATTCAGCAAGAGAGGATTTAAGA 541

Qy 1244 aaagactgtatcttcacattgatccatcaactgctcgacactgatgatccctcgcc 1303

Db 542 AAAGACTGTATCTTCAACATTTGACCCATCAACCCCGAGAGCTCGATGATGCCCTCTCC 601

Qy 1304 tgcaggcggtcactgatgacaccttcgaagtggcgctccacatcgccgatgtgattac 1363

Db 602 TGCAGCCACTCGCTGACGGCAACTTCAAGGTGGGAGTTTCATTTGCTGACGTGAGTATC 661

Qy 1364 ttgttctcaggagatcctcttggataaagttagctgtctgagagagccac-aagtgtcta 1422

Db 662 TTTGTTCCGGAGGATCTGATCTGGATAAAGTGGTGGCGAGAGGGCTACAAGGCTGTA 721

Qy 1423 ctgtgtccagaagtggtcccatctccagctctctgtgtgaggaactctgcagct 1482

Db 722 CTGTGTTCAAGGTGGTCCCATGCTTCCAGGC-TGTGTGTGAAGAGCTGTGCAGGCT 780

Qy 1483 caaccctactgacagactgacctctctctgtgatctggaagctgacccctgagggcaa 1542

Db 781 TAACCCAAATGTTGCAAAAGCTGACCTT-TCTGTGATTTGGGCACTGGGCCACAGAGGCAAA 839

Qy 1543 gatccttgaagagtgtttggcc 1565

Db 840 AACTTGAAAGGGTTGGCGGAC 862

RESULT 10

BI988256

LOCUS 3223-16 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence.

DEFINITION

ACCESSION BI988256

VERSION BI988256.1 GI:17959235

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 600)

AUTHORS Mu.X., Zhao.S., Pershad.R., Hsieh.T.-F., Scarpa.A., Wang.S.W., White.R.A., Beremand.P.D., Thomas.L.I., Gan.L. and Klein.W.H.

TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis

JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)

COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.

FEATURES Location/Qualifiers

source 1..600

BASE COUNT 143 a 171 c 149 g 137 t

ORIGIN

Query Match 18.9%; Score 586.4; DB 10; Length 600;  
Best Local Similarity 99.7%; Pred. No. 1.le-91;  
Matches 598; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1324 caccctcgaagtggcggtccacatcgccgatgtgagttactttgttcctgagggatcctc 1383

Db 1 CACCTTCGAAGTGGCGCTCCACATCGCGATGTGAGTTACTTTGTCTCGAGGGATCCTC 60

Qy 1384 ttgtgataaagtactgctgagagagccacaagtgtctacttgggtccagaaggtggtccc 1443

Db 61 TTTGGATAAAGTACTCTCTGAGAGAGCCACAAAGTGTCTACTTTGGTCCAGAGAGTGGTCCC 120

Qy 1444 catgtctccagggtctgtgtgaggaactctgcagcctcaaccccatgactgacaagct 1503

Db 121 CATGCTTCCAGGGCTTCTGTGTGAGGAACCTGCGAGCTCAACCCCATGACTGACAAAGCT 180

Qy 1504 gacctctctgtatcgaagctgacccctgaagcaagatccttgaagagtgtgttg 1563

Db 181 GACCTTCTCTGTATCTGAGAGCTGACCCCTGAAGGCAAGATCTTTGAAGAGTGGTTGG 240

Qy 1564 ccgcactatcctcgttcttgcaacaaactgagctacgaccatgcccagagatgatoga 1623

Db 241 CCGCACTATCATCGCTTCTTGCACCAAACTGAGCTAGCACCATGCCAGAGCATGATCGA 300

Qy 1624 aaatcacaactgagaagatccctgtaggaagagcttcccccaatttctccagagacagct 1683

Db 301 AAATCCAACTGAGAAGATCCCTGAGGAAGAGCTTCCCCCAATTTCTCCAGAGCAGCGT 360

Qy 1684 cgaggagtgcaccagcagctctgaaactgcacagcatgcaagcaactccgccca 1743

Db 361 CGAGGAGGTGCACCGACGACTCTGAACCTGCACAGCATTTGCAAGCAACTCCGCCGCCA 420

Qy 1744 gcgcttttagatggcgacactccgcttttagatcagctgaaagctgttttactctggacca 1803

Db 421 GCGCTTTGTAGATGGCGACTCCGCTTTAGATCAGCTGAAAGCTTGTCTTACTCTGGACCA 480

Qy 1804 tgagactgagctgcctcaagatgtcacatctatagta-ccgagacagacaagcttg 1862

Db 481 TGNAGACTGGATTGGCTCAAGGATGTACATCTATGAGTACCCGAGACGACACAGCTTG 540

Qy 1863 tagaggagtctatgctcctggccacatggcggtggcccaagatcttccgcaccttcc 1922

Db 541 TAGAGAGTTTATGCTCTCTCGCCCAACATGGCGGTGGCCCAAGATCTTCCCGCACCTTCC 600

RESULT 11

BI9804017

LOCUS 0244-55 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence.

DEFINITION

ACCESSION BI9804017

VERSION BI9804017.1 GI:17950931

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```

QY 1538 gcaagatccttgaagagtggttggccgcactatccgttcttgaccacaaactgaac 1597
Db 421 GCGAAGATCCTTGAAGAGTGGTTGGCCGCACTATCATCGTTCTTGGACCAAACTGAGC 480
QY 1598 tacgacatgccagagatgatcgaaataccaactgagaagatccctgaggaagcctt 1657
Db 481 TACGACCATGCCAGAGAGCATGATCGAAATCCAACCTGAGAAGATCCCTGAGGAAGAGCTT 540
QY 1658 ccccaatttctccagcacagcgtcgagg-aggtgcaccagcagctcctgaacctaca 1716
Db 541 CCCCCAATTCTCCAGAGCACAGCTCGAGGAAGGTGCACCAGCACTCTGAACCTG-A 599
QY 1717 cagcatcgaagcaactccgc 1739
Db 600 CAGCATTGCCAAGCAATCCGGC 622

RESULT 13
LOCUS BG825206 735 bp mRNA linear EST 22-MAY-2001
DEFINITION 602747617F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4900492 5',
mRNA sequence.
ACCESSION BG825206
VERSION BG825206.1 GI:14172793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1794 row: a column: 05
High quality sequence stop: 732.
FEATURES
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4900492"
/collection="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOT87; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT 167 a 217 c 200 g 151 t
ORIGIN

Query Match 18.4%; Score 570.8; DB 10; Length 735;
Best Local Similarity 86.1%; Pred. No. 5.2e-89;
Matches 632; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
```

```

QY 1303 ctgacggcctcactgacgttcgaagtgaggcgcacacatcgccgatgagttta 1362
Db 2 CTCGAGCCACTCGCTGACGGCAACTTCAAAGTGGGAGTTCACATGTCACGTGAGTTA 61
```

```

QY 1363 ctttgtctcaggagatcctcttggataaagtagctgctgagagagcccaagtgtcta 1422
Db 62 CTTTGTTCGGAGGAGATGATCTGGATAAAGTGGCTGCCGAGAGGGCTACAAGCGCTTA 121
QY 1423 cttgggtccagaaggtgggtcccccattgtcccaaggtctctgtgtgaggaactctgagcct 1482
Db 122 CTTGGTTCAAAAGGTGGTTCCTCCATGCTTCCAGGGTGTGTGTGAGGAGCTGTGCAGCCT 181
QY 1483 caaccocatgactgacaagctgaccccttctctgtgattcgtggaagctgacccctgaagcaa 1542
Db 182 CAACCCCATGTCCGACAAAGCTGACCTCTCTGTGTGATCTGGACACTGACTCCAGAGGGCAA 241
QY 1543 gatccttgaagtggtttggccgcactatccgttcttggaccacaaactgagctaca 1602
Db 242 GATCCTTGATGAATGGTTTGGCCGGACCATCCCTGCTGCACCAAACTTAGCTACCA 301
QY 1603 ccatgccagagcatgatcgaataatcaactgagaagatccctgaggaagagcttcccc 1662
Db 302 GCATGCACAGCATGATTGAAGGCCCAACTGAGAAATCCCTCGAAAGAGAGCTGCCCC 361
QY 1663 aatttccagagcacagcgtcgagggtgcaccagcagctcctgaaacctgcacagcat 1722
Db 362 CATTTCCTCCAGAGCATAGCAGCGAGGAGGTACACCAGGCCCTCTTGAATCTCCACGAA 421
QY 1723 tgcaagcaactccgcgccagcgtttgtatagtcgcactcogtttagatcagctgaa 1782
Db 422 TGCCAAAGCAGTTACGCCAGCAGCGCTTGTGGACGGCGCACTTCGTTGTGATCAGCTAAA 481
QY 1783 gctgtcttactctggaccatgagactggactgacctcaagatgtcacatctatagta 1842
Db 482 GCTTGTCTTCACTCTGGACCAACGAGACCGGATTCCTCAAGGATGTCATATCTATGAGTA 541
QY 1843 Ccgagacacaaacagcttctagagagttctatctcctgcacacatcggtgagccca 1902
Db 542 CGCCGAGAGCAACAAAGCTCTGTGGAGAGTTCATGCTCTGTGGCCAAACATGAGTGGCCCA 601
QY 1903 caagtcttcgcacctccctcctgagcagcctcctgcgcgcgcacatcccccacacagac 1962
Db 602 CAAGATCCACGCGCCCTTCGCCGAGCAGGCCCTGCTGCCGCGGACCCGCCGCCCAAC 661
QY 1963 gaagtgtcagtgacctggtgagttctgtgaccagatggagctcccatgagatgacg 2022
Db 662 AAGGATGCTCAGTCACTGGTGGAAATTCGCAACCATGAGTGGGGCTGCCCTGTGACTTCAG 721
QY 2023 ctctcagggggccc 2036
Db 722 CTCCGCGAGGAGGCC 735

RESULT 14
LOCUS BG746662 727 bp mRNA linear EST 15-MAY-2001
DEFINITION 602703970F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857353 5',
mRNA sequence.
ACCESSION BG746662
VERSION BG746662.1 GI:14057315
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 727)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
```

http://image.llnl.gov  
plate: LLCM1709 row: k column: 18  
High quality sequence stop: 727.

FEATURES

```

1: ..72/
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4857353"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DHI0B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
158 a 243 c 194 g 132 t

```

```
Db 242 GAGGACTGCAATTTTCCTGGGCGAGCTGGCTAAGAGTCTTGGGCGAGCTGGTAAATT 301
QY 1106 gagcctgaacagaaaggatactgacagaatatgggtgagcttctctgattctcttca 1165
    |||||
Db 302 GAGCCTGAACAGAAAGGAACTACTAACAGAGTATGGCGTGGATTTCTGTGATTTCTCTCA 361
    |||||
QY 1166 gaagttcttgaatgtctccctcaaaagcctgccctgacaaatcccacctgatgaggtggc 1225
    |||||
Db 362 GAAGTTCTAGAAATGTTTCTCAAGGCCTGCCATGGACAATTCACACAGAGGAGTTCAGC 421
    |||||
QY 1226 aagagaagagacctaaagaaagactgtatcttcaccatgatccatcaactgctcgcgac 1285
    |||||
Db 422 AAGAGAGGGGATTTAAGAAAGAGACTGTATCTTCACCATTGACCCATCAACGCCCGCAGAC 481
    |||||
QY 1286 ctgcatgatgccctgcctgcagggcgtcactgatggcaccttogaagtggcggtccac 1345
    |||||
Db 482 CTCGATGATGCCCTCTCCTGCAAGCCACTCGCTGACGGCAACTTCAAAAGTGGGAGTTTAC 541
    |||||
QY 1346 atcgccgatgtgagttaacttttctgaggatcctcttggataaagtactgctgag 1405
    |||||
Db 542 ATTGCTGACGTGAGTTACTTTGTTCCGGAGGGATCTGATCTGGATCAAGTGGCTGACGAG 601
    |||||
QY 1406 agggccacaagtgtacttgggtccagaaggtgggtcccatgcttcccaggcttctgtgt 1465
    |||||
Db 602 AGGGCTACAAGCGTCTACTTGGTTTCACACGGTGGTCCCAATGGCTTCCAGGCTGCTGTGT 661
    |||||
QY 1466 gaggaactctgcagcctcaaccccatg 1492
    |||||
Db 662 GAAGAACCCTGTGCTCCTCAAAACCATG 688
```

Search completed: July 8, 2002, 08:38:54  
Job time: 2461 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 07:58:23 ; Search time 2322.19 Seconds  
(without alignments)  
18017.716 Million cell updates/sec

Title: US-09-647-377-8  
Perfect score: 3100  
Sequence: 1 cggcgccgcgcgcctcccg.....tgaaaaaaaaaaaaaaaaaaaaa 3100

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description
1	597	19.3	725	10	BE380437	BE380437	601270510
2	578	18.6	1158	10	BI650394	BI650394	603296125
3	564	18.2	701	10	BG864079	BG864079	60297865
4	558	18.0	695	10	BI078196	BI078196	602872520
5	534	17.2	682	10	BI661242	BI661242	603304587
6	530	17.1	674	9	BB037094	BB037094	BB037094
7	527	17.0	681	10	BF450541	BF450541	uz67ff12.y
8	490	15.8	600	10	BI988256	BI988256	3223-16 M
9	490	15.8	684	9	BB636311	BB636311	BB636311
10	490	15.8	692	9	BB615143	BB615143	BB615143
11	479	15.5	521	10	BE534295	BE534295	601233534
12	470	15.2	506	10	BE910893	BE910893	601661970
13	451	14.5	622	10	BE284271	BE284271	601087339
14	449	14.5	451	9	AI597514	AI597514	vj50e11.x
15	443	14.3	444	10	BE380350	BE380350	601270414
16	412	13.3	417	9	BB751180	BB751180	BB751180
17	412	13.3	425	9	AI585294	AI585294	vj50e11.y

18	382	12.3	494	9	AA756851	AA756851	vu20c02.i
19	379	12.2	439	9	BB821152	BB821152	BB821152
20	370	11.9	600	10	BG804017	BG804017	0244-55 M
21	334	10.8	529	12	A2937324	A2937324	2M0195P24
22	328	10.6	427	9	AI464025	AI464025	vw72a02.y
23	317	10.2	455	10	BI645707	BI645707	603275018
24	308	9.9	643	9	BB620465	BB620465	BB620465
25	308	9.9	898	10	BF182212	BF182212	601804759
26	292	9.4	416	9	AI322478	AI322478	mh19c11.y
27	290	9.4	468	9	BB822910	BB822910	BB822910
28	289	9.3	427	9	AA014141	AA014141	mh19c11.i
29	288	9.3	439	9	AA575234	AA575234	vh19b05.i
30	284	9.2	651	9	BB636062	BB636062	BB636062
31	248	8.0	301	9	BB654969	BB654969	BB654969
32	245	7.9	301	9	BB268254	BB268254	BB268254
33	245	7.9	430	9	AI324115	AI324115	mh19c11.x
34	242	7.8	529	9	BB771440	BB771440	BB771440
35	232	7.5	509	12	A2582558	A2582558	1M0376J03
36	229	7.4	659	9	AV262261	AV262261	AV262261
37	223	7.2	405	9	AI558033	AI558033	vw72a02.x
38	212	6.8	239	9	AW123728	AW123728	UI-M-BH2.
39	212	6.8	239	10	BE952835	BE952835	UI-M-BH4.
40	208	6.7	377	9	BB793031	BB793031	BB793031
41	188	6.1	290	9	BB439487	BB439487	BB439487
42	184	5.9	468	12	A2401686	A2401686	1M0168N03
43	180	5.8	332	9	AA855380	AA855380	vw72a02.i
44	178	5.7	456	9	AA553101	AA553101	vj50e11.i
45	166	5.4	193	10	BE648481	BE648481	UI-M-BH2.

ALIGNMENTS

RESULT 1  
BE380437 601270510F1 NCI\_CGAP\_Maml Mus musculus cdna clone IMAGE:3599570 5',  
LOCUS mRNA sequence.  
DEFINITION 725 bp mRNA linear EST 21-JUL-2000  
ACCESSION BE380437  
VERSION BE380437.1 GI:9325802  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 725)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM8782 row: d column: 03  
High quality sequence stop: 596.

FEATURES  
Location/Qualifiers  
1..725  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3599570"  
/clone\_lib="NCI\_CGAP\_Maml"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator

```
BASE COUNT      206 a 172 c 179 g 167 t 1 others
ORIGIN

Query Match      19.3%; Score 597; DB 10; Length 725;
Best Local Similarity 100.0%; Pred. No. 5.1e-160;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 560 tttgacatcacccctcgagcagtcctcggaagagctgagtgctcctgattcattata 619
Db 1 TGTGGACATCACCCCTGCGAGCTCCCGAAGAGCTGGAGTGCTGCTGATGTCATTATA 60
QY 620 gagctcagttgatgacagagactcgaagatagacatggaacaccagctggctggtt 679
Db 61 GAGGCTCAGTTTGTATGACAGGCTCAAGAGATAGACATGGCAACACCAAGTGGCTGTT 120
QY 680 gatgtttaaagaattgcaatctactcctgacagaggaagaagattctagtact 739
Db 121 GATGGTGTAAAGAAATGTCAATCTCTACTCTGACAGAGGAAGAAAGATTCAGTACT 180
QY 740 ccagttatgaagatgagaacacacccccataccacagggacacagaggtttatcagaag 799
Db 181 CCAGTTATGAAGATGAGAACACACCCCATACACAGGACACAGAGGCTTATCAGAGAAG 240
QY 800 tcacttcaagaattcagcaaaagtgtttatctcttgagaaaaagcattctcgagcagca 859
Db 241 TCACCTCAGAAATCAGCAAGGTGTACATCTTGAGAAAAACATCTCGAGCAGCA 300
QY 860 actggaatcctgaactcttgctgataagaacagtgacctgtttaagaatacgcctg 919
Db 301 ACTGGATCCTGAACTCTGGCTGATAAAGAACAGTGACCTGTTTAAAGAAATAGCCCTG 360
QY 920 tttctccttcagacacccgagtagctagatttacgtacctcctcaagagctgtcccccag 979
Db 361 TTTTCTCCTTCAGACCACCGAGTACCTAGAAATTTACGTACCTCTCAGGACTGTCCCCAG 420
QY 980 gacttcataccagccttaagacttgccaaacacgcgtcttcactgcgcgcaatagac 1039
Db 421 GACTTCATGACCCGACCTTAAGAGCTTTGCCAACACGCCTGTTTCATCTGCCGATCATAGAC 480
QY 1040 tgaagaggagctgtaattttgcccctggggcaactgctaaagactcttggcagagctgg 1099
Db 481 TGAAGAGGAGCTGTAAATTTTGCCTGGGGCACTGCGTAAAGAGCTGTGGCAGGCTGGT 540
QY 1100 gaatacagcctgaacagaaggatatactgacagaatatgctggtggaactctctgat 1156
Db 541 GAATCGACCTGAACAGAGGGATACTGACAGAAATATGGTGTGAGCTTCTCTGAT 597

RESULT 2
BI650394
LOCUS      603296125F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5336719 5',
DEFINITION mRNA sequence.
ACCESSION BI650394
VERSION    BI650394.1 GI:15564630
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 1158)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-x@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL).
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
```

```
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11855 row: a column: 08
High quality sequence stop: 762.
Location/Qualifiers
1..1158
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5336719"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/Note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NCI_CGAP Library."
BASE COUNT      351 a 259 c 326 g 221 t 1 others
ORIGIN

Query Match      18.6%; Score 578; DB 10; Length 1158;
Best Local Similarity 100.0%; Pred. No. 8.3e-155;
Matches 578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 gagcgagcctcgacaactgagctgctggaaggcagaggaactctgagctgaatagtag 80
Db 27 GAGCGACGCTCGTGACAACTGAGCTGCTGAAGGCAGGAGGAACTCTGAGCTGAATAGTAG 86
QY 81 tgggtcccttaactctgagaagaagacgcacctggaaaccagtaataaacaccatcgtact 140
Db 87 TGGGTCCCTGAACTCGGAGAGAGAGCGCCCTTGGAAACAGTAATGAACCATCTGACT 146
QY 141 acaagctgaacctcgctccggggaccccccagagtggtcctcctgctggttgcccca 200
Db 147 ACAAGCTGAACCTTCGCTCGCGGGACCCCGAGAGGTGTCTCTGTGTGGCCCGA 206
QY 201 gtgctgttggtcctcgccaggtgacaaaagtcaaaagaacagtcacatgcagagaaga 260
Db 207 GTGCTGTGTGCTGTGCCAGGTGACAAAAGTCAAAAGATCCATGCCGAGGAAGA 266
QY 261 aaagagacattattgaaacctacatgtccaaggaggatgttcagaagcttgaagagag 320
Db 267 AAAAGACATATTGTAACACTACATGTCCAAGGAGAGATTTTCAGAAGGCTTTGAAGAGAG 326
QY 321 gaacacttaccagggtgtattgagaatcaaaccccaagaagtcttcataagccttcattc 380
Db 327 GAACACTTATCCAGGCTGTATTGAGAAATCAACCCCAAGAAAGTTCATGAGCCITCATTC 386
QY 381 cttctccggatgtgctatcgggacattttattgatggagtggtgtcgtogtaataagacct 440
Db 387 CTTCTCCGGATGTGTATCGGGACATTTTATTGATGAGTGTGTGCTCGTAATAGAGCT 446
QY 441 taaatggggagacctgtggtgttaaaactgctcctcagagatcagtggaagcagcttaaac 500
Db 447 TAATGGGACCTTGTGTTGTTAAACCTGCTCTCAGGATCAGTGAGGAGCGAGTTAAAC 506
QY 501 cagagacaaatgacaagaataatagaagctacttatgaagctgacatccctgaagagggct 560
Db 507 CAGAGCAATGACAAAGAAATAGAAAGCTACTTATGAAGCTGACATCCCTGAAGAGGCT 566
QY 561 gtgacatcacccctcgagcagctcccgaaagcgtgg 598
Db 567 GTGGACATCACCCCTGACAGAGTCCCGGAAGGCTGG 604

RESULT 3
BG864079
LOCUS      602797865F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4918576 5',
DEFINITION mRNA sequence.
```

ACCESSION	BG864079
VERSION	BG864079.1 GI:142114617
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 701)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: coaphs-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov. Plate: LLAM10832 row: b column: 17 High quality sequence stop: 701.
FEATURES	Location/Qualifiers
source	1..701
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/clone="IMAGE:4918576"
	/clone_lib="NIH_CGAP_Mam4"
	/tissue_type="tumor, gross tissue"
	/lab_host="DH10B"
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NIH_CGAP Library."
BASE COUNT	171 a 194 c 174 g 162 t
ORIGIN	
Query Match	18.2%; Score 564; DB 10; Length 701;
Best Local Similarity	99.8%; Pred. No. 1.2e-150;
Matches 614; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Qy	1105 cgagctgaacacagaaggatgactgcagacaatgtgtggacctctctgattcttc 1164
Db	1 CGAGCCTGAACACAGAAGGGATGACTGCAGAAATATGTGTGGACTTCTGATTCTCTTC 60
Qy	1165 agaagtcttgatgtctccctcaaacctgcctcgccctggacaatccccactgatgaggtggg 1224
Db	61 AGAAGTCTTGTAATGTCCTCCCTCAAGGCCCTGCCTGGACAATCCCACCCTGATGAGGTGG 120
Qy	1225 caagagaagaccctaagaagactgtattcttcacattgacatcatcaactgctcgcgga 1284
Db	121 CAGAGACAGACCTTAAGAAAGACTGTATCTTCACCATTTGATCCATCAACTGCTCGGGA 180
Qy	1285 ccttgatgatgccctgcctgcagcggcctcaactgatggcaccttcgaagtgccgtcca 1344
Db	181 CCTTGATGATGCCCTCGCTGCAGCGGCTCACATGTATGGACATTCGAAGTGGGGTCCA 240
Qy	1345 catcgccgatgtgagttaactttgtcctgaggatcctctttggataaaagttagctgtga 1404
Db	241 CATGCCGATGTGAGTACTTTGTTCCTGAGGGATCCTCTTTGGATAAAGTAGCTGTGA 300
Qy	1405 gagagccacaagtgtctacttgggtccagaaggtgggtcccactgcttcccaggcttctgtg 1464
Db	301 GAGAGCCACAAGTGTCTACTTGTGTCCAGAAAGGTGGTCCCATGCTTCCCAGAGCTTCTGTG 360
Qy	1465 tgaggaaactctgcagctcaaccccaagactgacagcaagctgaccttctctgtgacttgaaa 1524

Db	361 TGAGGAACCTGCGAGCCTCAAGCCCATGACTGACAAGCTGACCTTCTCTGTGATCTGGAA	420
Qy	1525 gctgaccttcgaaggcaagatccttgaagagtgggtttggccgcaactatcatcgtttcttg	1584
Db	421 GCTGACCCCTGAAGGCAAGATCCTTGAAGAGTGGTTTTGGCGGCACTATCATCGTTCTTG	480
Qy	1585 caccaaatgagctagcaccatgccacagagcatgacatgcacaaatcccaactgagaagatccc	1644
Db	481 CACCAAACCTGAGCTAGCACCATGCCAGAGCATGATGAAAATCCAACCTGAGAAGATCCC	540
Qy	1645 tgaggaagagcttcccccatttctccagagcacagcgtcgaagaggtgcaccagagcagt	1704
Db	541 TGGAGGAAGAGCTTCCCCCAATTTCTCCAGAGCACAGCGTCGAGAGGTGCACAGGCAGT	600
Qy	1705 cctgaacctgcacag 1719	
Db	601 CCGTAACCTGCACAG 615	
RESULT	4	
BIOTR196	602872520F1 NCI_CGAP_Mam2 Mus musculus cdna clone IMAGE:5004341 5'	EST 20-JUN-2001
LOCUS	mRNA sequence.	
DEFINITION	BIOTR196	
ACCESSION	BIOTR196.1 GI:14496526	
VERSION	BIOTR196	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 695)	
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: coaphs-remail.nih.gov	
COMMENT	Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1042 row: h column: 06 High quality sequence stop: 689.	
FEATURES	Location/Qualifiers	
source	1..695	
	/organism="Mus musculus"	
	/strain="FVB/N-3"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:5004341"	
	/clone_lib="NCI_CGAP_Mam2"	
	/tissue_type="tumor, biopsy sample"	
	/dev_stage="5 months"	
	/lab_host="DH10B"	
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"	
BASE COUNT	191 a 151 c 200 g 153 t	
ORIGIN		
Query Match	18.0%; Score 558; DB 10; Length 695;	
Best Local Similarity	99.9%; Pred. No. 6e-149;	
Matches 678; Conservative	0; Mismatches 0; Indels 1; Gaps 1;	
Qy	10 cggcctccggagagcagcctgtgacacactgactgctgaagcagaggaactctgag 69	
Db	11 CGGCTCCCGGAGGAGCGCTCGTGACAACTGAGCTGCTGAAGGAGGAGGAACCTCTGAG 70	
Qy	70 ctgeatagtagtggtccctgaatctggagagaagacgcacaccttggaaaccagtaatgaa	129



**AUTHORS**  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Sasaki, Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M., and Hayashizaki, Y.

**TITLE**  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

**JOURNAL**  
Unpublished (2001)

**COMMENT**  
On Jun 10, 2000 this sequence version replaced gi-8443480.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y., and Hayashizaki, Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

**FEATURES**  
source  
1..674  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="5930423H14"  
/clone\_lib="RIKEN full-length enriched, 13 days embryo forelimb"  
/sex="mixed"  
/tissue\_type="forelimb"  
/dev\_stage="13 days embryo"  
/lab\_host="Dh10B"  
/notes="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCTTCAGATTAATAATTAATCCCCCCCCCC 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adaptor of

sequence [5' GAGAGAGAGATCTTCAGATTAATAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 149 a 166 c 188 g 171 t  
ORIGIN

Query Match 17.1%; Score 530; DB 9; Length 674;  
Best Local Similarity 100.0%; Pred. No. 5.4e-141;  
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2555 aagccagagactcattctgtttggaagcctgatgacctgaagagagagcacaacagcag 2614  
Db 142 AACCCAGAGCTCACCTCTGTTGGAGCCCTGATGACCTTGAAGAGGAGCCACACAGCAG 201

QY 2615 gtcatcacattctcagcctgtgtgattgtctcgcagcagagagccacagcctcaag 2674  
Db 202 GTCATCACCATCTTCAGCCTGTGTGATGTGTCTTCAGCGCAGAGGCCACAGCCCTCAAG 261

QY 2675 tacagtctatcctgaacgaccagcctgagaaaggcgtctgatgagagagcctgaggac 2734  
Db 262 TACAGTGCCTATCTCTGAACGACACAGGCCCTGGAGAGGCCCTCTGATGAGAGGCCCTGAGGAC 321

QY 2735 tgaatgctagcccaagccagcctgtgtcctgcctaccctcctgctgcttttaggaatagga 2794  
Db 322 TGAATGCTAGCCCAAGCCAGCCCTGTGCTGCCTTACCCTGCTGCTTTTGAAGTAAGGA 381

QY 2795 ccttttgacacaaagggagatttaatttggtttttaaacactcaggggttttttta 2854  
Db 382 CCTTTTGACACAAAGGGGATTTTAAATTTGGTTTTTAACAACCTCAGGGGTTTGTTTTAA 441

QY 2855 tttttattttctctttatttttacttttgagctcagtttttaaatgaactggaaggtta 2914  
Db 442 TTTTATTTTCTCTTTTATTTTACTTTTGCAGCTCAGTTTTTAAATGAACCTGAAGGTTA 501

QY 2915 ggggtcagggcagggatgctgagcctgagcctgcttccctgagcagagagatccca 2974  
Db 502 GGGGTACAGGCGAGGGGATCTCTAGGCCCTGGCCTGTGCTTCCCTGAGCAGAGAGATCCCA 561

QY 2975 gtccctctggcagggcagcccgcttctacagggcagcccaactgcccttccctcccaag 3034  
Db 562 GTCTCTCTGGCAGGCGAGCCCGCTTCTACCGCGCAGCCACTGCTCTCCTGCTCCAGG 621

QY 3035 aaatgggggtttcagcaaatcagtgctgaatgaataaatcaagtgtaa 3084  
Db 622 AAATGGGGGGTTTTCAGCAAAATCAGTGTCTATGGAAATAAAATCAAGTGTGAA 671

RESULT 7  
BF450541  
LOCUS  
DEFINITION  
us67fl2.y1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3674159 5' similar to SW:YR86\_CAEEL Q09568 HYPOTHEICAL 93.7 KDA PROTEIN F4E8.6 IN CHROMOSOME III. [2] SW:YR86\_CAEEL ; mRNA sequence.  
BF450541  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF450541 681 bp mRNA linear EST 29-DEC-2000  
us67fl2.y1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3674159 5' similar to SW:YR86\_CAEEL Q09568 HYPOTHEICAL 93.7 KDA PROTEIN F4E8.6 IN CHROMOSOME III. [2] SW:YR86\_CAEEL ; mRNA sequence.  
BF450541  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF450541.1 GI:11516710  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 681)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapsu@mail.nih.gov](mailto:cgapsu@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/URL at:  
image.llnl.gov/image/html/iresources.shtml

MGI:1434927

Seq primer: -40RP from Gibco  
High quality sequence stop: 440.

## FEATURES

Location/Qualifiers

source

1..681  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:3674159"  
/clone\_lib="NCI-CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DHL08"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT

160 a 206 c 169 g 144 t 2 others

ORIGIN

Query Match 17.0%; Score 527; DB 10; Length 681;  
Best Local Similarity 99.8%; Pred. No. 3.8e-140;  
Matches 577; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1589 aaactgagctagaccatgccagagcatgatcgaaatccaactgagaagatccctgag 1648

Db 1 AAACCTGAGCTAGGACGATGCCAGAGCATGATGAAATCCAACTGAGAAGATCCCTGAG 60

QY 1649 gaagagcttcccccaatttccagagcacagctcgagaggtgcacacaggaactctg 1708

Db 61 GAGAGCTTCCCCCAATTTCTCCAGACACAGCGTCCGAGAGGTGCACAGGAGTCCGT 120

QY 1709 aactgcacagcattgcaagaactccgcgcagcgctttgttagatggcgactccgt 1768

Db 121 AACCTGCACAGCATTTGCCAAGCAACTCCGCCGCGAGCGTTTGTAGATGGCGCATCCGT 180

QY 1769 ttgatcagctgaagcttgctttactctgaccatgagactggaactgcctcaagatgt 1828

Db 181 TTAGATCAGCTGAAGCTTGCTTTTACTCTGGACCATGAGACTGGATTGCCCTCAAGATGT 240

QY 1829 cacatctatgdiaccgagacagcaacaagctgtgagggagtctatgctctctggccaac 1888

Db 241 CACATCTATGATACCGAGACAGCAACAAGCTTGTAGAGGAGTTCTATGCTCTCTGGCCAAC 300

QY 1889 atggcggtggcccaagatcttcgcacacttccctgagcagggccctgtgcgcggcat 1948

Db 301 ATGGCGGTGGCCCAAGATCTCCGCACCTTCCCTGAGCAGGGCCCTGCTGCGCGGGCAT 360

QY 1949 cccccaccagagaagaatgctcagtgacctgggtggagttctgtgaccagatggggctg 2008

Db 361 CCCCCACCAGAGAGAAGATGCTCAGTGAACCTGGTGGAGTTCTGTGACCAGATGGGGCTG 420

QY 2009 cccatgatgtcagctctgcagggccctaaataaaacccctgactaagaacattggagat 2068

Db 421 CCCATGGATGTACGCTCTGCAGGGGCCCTAAATAAAGCCCTGACTTAAGACATTTGGAGAT 480

QY 2069 gacaagtactcttggcccggaagggtgctcaccacaactgactccccggcccatgcaag 2128

Db 481 GACAAGTACTCTCTGGCCGCGAAGGAGGTGCTCACCAACATGTACTCCCGGCCCATGCGAG 540

QY 2129 atggcacgtacttctgctctggatgctgcagagacca 2166

Db 541 ATGGCACTGTACTTCTGCTCTGGATGCTGCAGGACCA 578

RESULT 8

BI988256

LOCUS

DEFINITION

3223-16 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,

mRNA sequence.

ACCESSION

BI988256

VERSION

BI988256.1 GI:17959235

KEYWORDS

EST

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 600)

AUTHORS

Mu X., Zhao S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,

White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.

TITLE

Gene expression in the developing mouse retina by EST sequencing

JOURNAL

Nucleic Acids Res. 29 (24), 4983-4993 (2001)

COMMENT

Contact: Klein WH

Department of Biochemistry and Molecular Biology

University of Texas M.D. Anderson Cancer Center

Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA

Tel: 713 792 3646

Fax: 713 790 0329

FEATURES

Location/Qualifiers

1..600

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

/tissue\_type="neural retina"

/dev\_stage="embryonic day 14.5 post-fertilization"

BASE COUNT

143 a 171 c 149 g 137 t

ORIGIN

Query Match 15.8%; Score 490; DB 10; Length 600;

Best Local Similarity 100.0%; Pred. No. 1.3e-129;

Matches 490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 caccttcgaagtggcgctccacatcgccgatgtgagttactttgtctcctgaagatcctc 1383

Db 1 CACCTTCGAGTGGGCGTCCACATCGCGATGTGAGTTACTTTGTTCCTTGAGGGATCCTC 60

QY 1384 ttgggataaagttagctgctgagagggccacagtgtctacttgggtccagaagtggtccc 1443

Db 61 TTTGGATAAAGTAGCTGCTGAGAGAGCCACAAAGTGTCTACTTGGTCCAGAGGTGGTCCC 120

QY 1444 catcttccacagcttctgtgaggaactctgcagctcaaccctgaactgaactgacagct 1503

Db 121 CATGCTTCCAGCGCTTCTGTGTGAGGAACCTCGAGCCTTCACCCCTCAGCTGACAGCT 180

QY 1504 gaccttctgtgactggaagctgacccctgaaggcaagatccttgaagagtggtttg 1563

Db 181 GACCTTCTGTGATCTGGAAGCTGACCCCTGAAGGCAAGATCCTTGAAGAGTGGTTGG 240

QY 1564 ccgcactatcatccgttctgcaccaactgaactgaactgaactgaactgaactgaactga 1623

Db 241 CCGCACTATCATCCGTTCTTGGCCAACTGAGCTAGCACCATGCCAGACATGATCGA 300

QY 1624 aaatccaaactgagaagatcccttgaggagagcttcccccaatttctccagagcagcgt 1683

Db 301 AAATCCAACTGAGAAGATCCCTTGAGGAAGAGCTTCCCCCAATTTCTCCAGAGCAGCGT 360

QY 1684 cgaagggtgcacagcgagctcctgaacctgcacagcattgcaaaagcaactccgcccaca 1743

Db 361 CGAGAGGTGCACCGGAGTCTTGAACCTGCACAGCATTTGCAAGAGCACTCCGCCGCCA 420

QY 1744 gcgcctttagatggcgcactcccttttagatcagctgaagcttgccttttactctggacca 1803

Db 421 GCGCTTTGATGTGGCGCACTCCGTTTAGATCAGCTGAAGCTTGTCTTTACTCTGGACCA 480

QY 1804 tgaactgga 1813

Db 481 TGAGACTGGA 490



Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tadawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,I.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)

# TITLE JOURNAL COMMENT

Contact: Yoshinide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watabiki,M., Itoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,  
Ishii,Y. and Hayashizaki,Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.  
Func. Genomics 2 prte, L72-L86 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

## FEATURES source

e mouse tissues.  
Location/Qualifiers  
1. .692  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="4930429A22"  
/clone\_lib="RIKEN full-length enriched, adult male testis"  
/sex="male"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/note="Site\_1: XhoI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAAGGATCCAGAGCTCTTTTITTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGCGCGCAATTAATTCAGTTAATAATTAATTCGCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites. "  
197 a 154 g 184 g 157 t

## BASE COUNT ORIGIN

Query Match 15.88; Score 490; DB 9; Length 692;  
Best Local Similarity 100.0%; Pred. NO. 1.2e-129;  
Matches 490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	cggcgccgcccgcctcccgaggagcgacgctcgtgacaaactgagctgctgaagcgagg	60
Db	18	CGCGCCGCGCGCTCCCGGAGCGAGCTCGTGAACACTGAGCTGTGAAGCAGGAGG	77
QY	61	aactctgagctgaa tagtagtggtccctgaa tctggagagaagacgccaccttgaacc	120
Db	78	AACTCTGAGCTGAATAGTAGTGGTCCCTGAACTCTGGAGAGAAGCCACCTTGAACC	137
QY	121	agtaatgaaccatcctgactacaagctgaaccttcggtctccggggaccctccagaggtg	180
Db	138	AGTAATGAACCATCTGACTACAAGCTGAACCTTCGGTCTCCGGGGACCCCGAGAGGTG	197
QY	181	gtcctctgtgtggtggcccgagtgctgtgtgtcctgcaggtgacaaaagtcaaaaga	240
Db	198	GTCTCTGTGTGTGGTGGCCGAGTGCTGTGTGCTCTTCGCCAGTCAACCAAGAA	257
QY	241	caagtcctatgcgaggaagaaaagacataatttgaacctacatgtccaaaggagatgt	300
Db	258	CAAGTCCATGCGAGGGAAGAAAGAGCATATTTGAACCTTACATGTCCAAGGAGGATGT	317
QY	301	ttcagaagccttgaagagaggaacacttatccaggggtgtattgagaatcaacccaaagaa	360
Db	318	TTCAGAAGCCTTGAAGAGAGAAACACTTATCCAGGTGTATTGAGAATCAACCAAGAA	377
QY	361	gtttcatgaaccttcattcttcctcggatgtgctcgggacatttttatgatggagt	420
Db	378	GTTCATGAAGCCCTCAITCTCTCCGATGGTGATCGGACATTTTATTGATGGAGT	437
QY	421	tgttctcgtataatgagccttaaatgggacaccttgtgtgtgttaaaactgcttcttgagga	480
Db	438	TGTTGCTGTAATAGAGCCCTAAATGGGACCTTGTTGTGTTGTTAAACTGCTCTCTGAGA	497
QY	481	tcagtggaag 490	
Db	498	TCAGTGGAG 507	

  

RESULT	11
BES34295	601235354Fl NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3599004 5',
LOCUS	BES34295 521 bp mRNA linear EST 09-AUG-2000
DEFINITION	mRNA sequence.
ACCESSION	BES34295
VERSION	BES34295.1 GI:9762940
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE	1 (bases 1 to 521)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Jeffrey Green M.D. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM8780 row: 1 column: 13 High quality sequence stop: 519. Location/Qualifiers 1. .521 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:3599004" /clone_lib="NCI_CGAP_Mam6" /sex="female, virgin" /tissue_type="infiltrating ductal carcinoma"

BASE COUNT 139 a 132 c 124 g 126 t  
ORIGIN

/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

Query Match 15.5%; Score 479; DB 10; Length 521;  
Best Local Similarity 100.0%; Pred. No. 1.9e-126;  
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 846 attctcgagcagcaactggcctcctgaaactcttgctgataagaacagtcgactgttta 905  
|||||  
Db 43 ATTCTCGAGCAGCAACTGGCATCTTGAACTCTTGCTGTAGAACAGAGTGACCTGTTTA 102  
|||||

QY 906 agaaatacgccctgttttctcctcagaccaccgagctacctaatttaagctactca 965  
|||||  
Db 103 AGAAATACGCCCTGTTTCTCTTCAGACACCGAGTACCTAGAAATTTAGTACCTCTCA 162  
|||||

QY 966 agagctgtcccccagactctcatgacccgacctaaagactttgccaacagctgttcatct 1025  
|||||  
Db 163 AGGACTGTCCCCAGGACTTTCATGACCCGACCTAAAGACTTTGCCCCAACGCTGTTCATCT 222  
|||||

QY 1026 gccgataatagatggaagagactgtaatttgcctggggcaactgactaagatc 1085  
|||||  
Db 223 GCCGATCATAGATGGAAGGAGACTGTAAATTTTGCCCTGGGCAACTGGCTAAGAGTC 282  
|||||

QY 1086 ttgggcagcgtggtgaaatcgagcctgaaacagaggggatactgacagaatatggtgg 1145  
|||||  
Db 283 TTGGCAGCGCTGGTGAATCGACTGAAACAGAGAGGATAC TGACAGAAATATGTTGTGG 342  
|||||

QY 1146 actctctgattctctctcagaagcttctgaaatgctccctcaagcctccctcgacaa 1205  
|||||  
Db 343 ACTTCTCTGATTTCTCTCAGAAGTCTTGAAATGTCCTCCCTCAAGGCTGCCCTGACAA 402  
|||||

QY 1206 tccacctgatgagtggtgggcaagagagactgaagaaagactgtattctcaccattg 1265  
|||||  
Db 403 TCCACCTGATGAGTGGGCAAGAGAGAGACTTAAGAAAGACTGTATCTTCCACCATG 462  
|||||

QY 1266 atccataactctcgacacttgatgacccctgcctcgagcggctcactgatggc 1324  
|||||  
Db 463 ATCCATCACTGCTCGGACCTTGATGATGCTCCCTGCGCTGAGCGGCTCAGTGATGGC 521  
|||||

RESULT 12  
BE910893  
LOCUS 601661970F1 NCI\_CGAP\_Mam1 Mus musculus cdna clone IMAGE:3962249 5',  
DEFINITION mRNA sequence.  
ACCESSION BE910893  
VERSION BE910893.1 GI:10407877  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 506)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM9128 row: k column: 18

High quality sequence stop: 506.  
Location/Qualifiers  
1..506  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3962249"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 136 a 105 c 145 g 120 t  
ORIGIN

Query Match 15.2%; Score 470; DB 10; Length 506;  
Best Local Similarity 100.0%; Pred. No. 7.1e-124;  
Matches 470; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 gagcgagcgtcgtgacaactgagctgctgaaggcaggaggaactcgtgagctgaatagtag 80  
|||||  
Db 1 GAGCGAGCGCTCGTGACAACTGAGCTGCTGAAGCGAGGAACTCTGAGCTGAATAGTAG 60  
|||||

QY 81 tgggtccctgaaatctggagagaagacgccaccttggaaaccagtaataagcattcctgact 140  
|||||  
Db 61 TGGGTCCCTGAATCTGGAGAGAGAGAGGCCACCTTGGANCCAGTAATGAACCATCTGACT 120  
|||||

QY 141 acaagctgaacctcgttcctcggggacccccagaggtgtgctcctctgtggtggccgca 200  
|||||  
Db 121 ACAAGCTGAACCTTCGGTCTCGGGGACCCCGAGAGTGTGCTCTGTGTGTGGCCCGA 180  
|||||

QY 201 gtgctgtgtgctcgcaggtgacaaaagtcataagaaagtcattgcgaggggaaga 260  
|||||  
Db 181 GTGCTGTGTGTGCTCGCCAGGTGACAAAAGTCAAAAGTCCATGCGAGGGAAGA 240  
|||||

QY 261 aaaagagcatatttgaacctacatgtccaaaggagagatgtttcagaaggcttggaagag 320  
|||||  
Db 241 AAAAGAGCATATTTGAAACCTACATGTCCAAGGAGGATGTTTCAGAAAGCTTGAAGAGAG 300  
|||||

QY 321 gaacactatccagggtgtattgagaatacaaccccaagagtttcatgaagccttcattc 380  
|||||  
Db 301 GAACACTATTCAGAGGTGATTGAGAATCAACCCAAAGAGTTTTCATGAAGCCTTCATTC 360  
|||||

QY 381 ctctccgagtggtgatcgggacattttattatgaggtgtgtgtcgtgaatagagcct 440  
|||||  
Db 361 CTCTCCGGATGAGTATCGGGACATTTTATTGATGAGGTGTGCTCGTAATAGAGCCT 420  
|||||

QY 441 taaatggggacctgtgtgtgtataaacctgctcctcctgaggtcagtggaag 490  
|||||  
Db 421 TAAATGGGACCTGTGTGTGTGTAATAACTGCTTCTCTGAGGATCAGTGGAAG 470  
|||||

RESULT 13  
BE284271  
LOCUS 601087339F1 NCI\_CGAP\_Mam5 Mus musculus cdna clone IMAGE:3482313 5',  
DEFINITION mRNA sequence.  
ACCESSION BE284271  
VERSION BE284271.1 GI:9161093  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 622)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.



QY 203 gctgttggtcttcgacaggtgacaaaagtcgaagcaagtcacatgcagggagagaaa 262  
|||||  
Db 183 gctgttggtcttcgacaggtgacaaaagtcgaagcaagtcacatgcagggagagaaa 242  
|||||  
QY 263 aagagcatattgaaacctacatgtcccaagagagatgttcagaagccttgagagagga 322  
|||||  
Db 243 AAGAGCATATTTGAACACCTACATGTCGAAGGAGGATGTTTCAGAGGCTTGAAGAGAGGA 302  
|||||  
QY 323 acattatccagggtgtatgagaatcaacccaagaagtttcataagcccttcattct 382  
|||||  
Db 303 ACATTATCCAGGGGTATTTGAGAATCAACCCAAAGAGTTTCATGAAGCCTTCATCCT 362  
|||||  
QY 383 tctccgagtgatcgagacatctttttattgatgagttgtctcgttaataagaccctta 442  
|||||  
Db 363 TCTCCGGATGATCGGACATATTTATTGATGAGTGTGCTCGTATAGAGCCCTTA 422  
|||||  
QY 443 aatgggacacctgtgtgtataaactgct 471  
|||||  
Db 423 AATGGGACCTTGTGGTTGTAAACTGCT 451  
|||||

RESULT 15  
BE380350  
LOCUS BE380350 444 bp mRNA linear EST 21-JUL-2000  
DEFINITION 601270414F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3599763 5',  
mRNA sequence.  
ACCESSION BE380350  
VERSION BE380350.1 GI:9325715  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 444)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM8782 row: 1 column: 04  
High quality sequence stop: 442.

FEATURES  
Location/Qualifiers  
1..444  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3599763"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 124 a 108 c 105 g 107 t  
ORIGIN  
  
Query Match 14.3%; Score 443; DB 10; Length 444;  
Best Local Similarity 100.0%; Pred. No. 3.6e-116;  
Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 802 acttcagaatacagaagtggtttacatcttggagaaaaagcattctcgcagcagaac 861  
|||||  
Db 1 ACTTCAGAAATCAGCAAGGTGTTTACATCTTGGAGAAAAAGCATTTCTCGAGCAGCAAC 60  
|||||

QY 862 tggcatctgaaactctcttgctgataagaacagtgacctgtttaagaatacgcctgtt 921  
|||||  
Db 61 TGGCATCTCTGAAACCTCTTTGGCTGATAGAAACAGTAGTACCTGTTTAAAGAAATACGCCCTGTT 120  
|||||  
QY 922 ttctcttcagaccaccaggtacctaagaatttacctctcctcctcaagagactgtccccagga 981  
|||||  
Db 121 TTCTCTCTCAGACCAACCGAGTACCTAGATTACGTACTCTCAAGGACTGTCCCGCAGGA 180  
|||||  
QY 982 ctctcagaccgcacctaagaactttgccacaacagctgtttcatctgcgcacatcatacactg 1041  
|||||  
Db 181 CTTCATGACCCCGACCTAAAGACTTTGCCAACACGCTGTTTCATCTGCCCGCATCATAGACTG 240  
|||||  
QY 1042 gaaggaggactgtaattttggccctggggcaactggcctaagagctcttggcaggctggtga 1101  
|||||  
Db 241 GAAGGAGGACTCTAATTTTGGCCCTGGGGCAACTGGCTAAGAGTCTTTGGGAGGCTGGTGA 300  
|||||  
QY 1102 aatcgagcctgaaacagaaggatactgacagaataatagtgtagacttctcatttctc 1161  
|||||  
Db 301 AATCGAGCCTGAACAGAGGGGATCTGCAGAAATATGGTGTGGACTTCTCTGATTTCTC 360  
|||||  
QY 1162 ttcagaagtcttgtaattgtctctccctcaagcctgcctggacaatccccacctgatgaggt 1221  
|||||  
Db 361 TTCAGAAAGTTCTTGAATGCTCTCCCTCAAAGCCTGCCTTGGCAATCCACCTGATGAGGT 420  
|||||  
QY 1222 gggcaagagaagagacctaaaga 1244  
|||||  
Db 421 GGGCAAGAGAAGAGACCTAAGGA 443  
|||||

Search completed: July 8, 2002, 10:26:46  
Job time: 8903 sec

---

OM of: US-09-647-377-9 to: EST:\* out\_format : pfs

Date: Jul 8, 2002 1:24 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09647377/runat\_08072002\_080020\_23053/app\_query.fasta\_1.940  
-DB=EST -OPMT=fastap -SUFFIX=olip2n.rst -GAPOP=4.500  
-CAPEXT=0.050 -MINMATCH=0.100 -LOOPCI=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -CAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=60.000 -XGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS-human40.cdi -LIST=45 -DLOCALIGN=200 -THR\_SCORE=quality  
-THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09647377 @CGN1\_1.7845 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-647-377-9  
Query length: 870  
Database: EST:\*  
Database sequences: 13736207  
Database length: -1841457050  
Search time (sec): 1636.090000

WARN: XGAPOP and XGAPEXT must be equal. Assuming XGAPOP=60.000  
WARN: XGAPEXT and XGAPEXT must be equal. Assuming XGAPEXT=60.000

score\_list:

sequence	Strd	Orig	zscore	Escore	Len	Documentation
gb_est2:BE380437	+	199.00	3689.37	3.0e-196	725	! BE380437 601270510F1 NCI_CGAP_M
gb_est2:BF450541	+	195.00	3615.07	4.1e-192	681	! BF450541 uz67f12.y1 NCI_CGAP_M
gb_est2:BI988256	+	186.00	3447.80	8.5e-183	600	! BI988256 3223-16 Mouse E14.5 re
gb_est2:BI6161242	+	178.00	3297.30	2.1e-174	682	! BI6161242 603304587F1 NIH_CGAP_M
gb_est2:BI078196	+	175.00	3241.08	2.8e-171	695	! BI078196 602872520F1 NCI_CGAP_M
gb_est2:BE534295	+	159.00	2944.19	9.6e-155	521	! BE534295 601235354F1 NCI_CGAP_M
gb_est2:BE534295	+	158.00	2919.45	2.3e-153	1158	! BE530394 603296125F1 NIH_CGAP_M
gb_est2:BE380350	+	147.00	2721.09	2.6e-142	444	! BE380350 601270414F1 NCI_CGAP_M
gb_est1:AA756851	+	127.00	2346.44	1.9e-121	494	! AA756851 vu20c02.r1 Barstead mc
gb_est2:BG864079	+	126.00	2325.10	2.9e-120	701	! BG864079 602978765F1 NIH_CGAP_M
gb_est2:BG10893	+	122.00	2252.80	3.1e-116	506	! BG10893 601661970F1 NCI_CGAP_M
gb_est1:BB636111	+	122.00	2250.43	4.2e-116	684	! BB636111 BB636311 RIKEN full-1e
gb_est1:BB636111	+	115.00	2122.83	5.4e-109	451	! BB615143 BB615143 RIKEN full-1e
gb_est2:BG804017	+	109.00	2008.51	1.3e-102	600	! BG804017 0244-55 Mouse E14.5 re
gb_est1:AA575234	+	106.00	1954.80	1.2e-99	439	! AA575234 vh19b05.r1 Soares mam
gb_est2:BF182212	+	102.00	1874.62	3.6e-95	898	! BF182212 601804759F1 NCI_CGAP_M
gb_est2:BE284271	+	99.00	1821.32	3.3e-92	622	! BE284271 601087339F1 NCI_CGAP_M
gb_est1:AI322478	+	97.00	1786.98	2.7e-90	416	! AI322478 mh19c11.y1 Soares mous
gb_est2:BF548574	+	97.00	1785.85	3.2e-90	483	! BF548574 UI-R-A0-at-d-12-0-UI-R
gb_est1:AA014141	+	96.00	1768.09	3.1e-89	427	! AA014141 mh19c11.r1 Soares mous
gb_est1:AI558033	+	85.00	1562.88	8.3e-78	405	! AI558033 vw72a02.x1 Stratagene
gb_est1:AI324115	+	81.00	1487.66	1.3e-73	430	! AI324115 mh19c11.x1 Soares mous
gb_gss:AA582558	+	77.00	1411.61	2.2e-69	509	! AZ582558 LM0376303F Mouse 10kb
gb_est1:AV262261	+	77.00	1409.66	2.8e-69	559	! AV262261 AV262261 RIKEN full-1e
gb_est1:AI904514	+	76.00	1390.92	3.2e-68	663	! AI904514 PM-BT057-100299-382 BT
gb_est2:BF979456	+	73.00	1330.94	6.9e-65	1110	! BF979456 602288005F1 NIH_MGC_9
gb_est1:BB620465	+	69.00	1260.31	5.9e-61	643	! BB620465 BB620465 RIKEN full-1e
gb_est1:BB037094	+	67.00	1222.57	7.5e-59	674	! BB037094 BB037094 RIKEN full-1e
gb_est1:BB636062	+	66.00	1204.14	8.0e-58	651	! BB636062 BB636062 RIKEN full-1e
gb_est2:BB684505	+	65.00	1186.52	7.7e-57	565	! BB684505 185915 MARC 4BOV Bos t
gb_est2:BB684505	+	61.00	1113.17	9.3e-53	468	! AZA01686 LM0168N03R Mouse 10kb
gb_gss:AA787575	+	60.00	1093.46	1.2e-51	536	! AW785775 117208 MARC 1PTG Sus s
gb_est1:AW787575	+	59.00	1074.86	1.3e-50	529	! BB771440 BB771440 RIKEN full-1e
gb_est2:BI344672	+	54.00	986.49	1.1e-45	270	! BI344672 373199 MARC 2PTG Sus s
gb_est2:BF405228	+	51.00	931.06	1.3e-42	248	! BF405228 UI-R-CAI-bis-e-02-0-UI
gb_est2:BE791331	+	51.00	921.41	4.5e-42	887	! BE791331 601582710F1 NIH_MGC_7
gb_est2:BG106547	+	51.00	919.38	5.8e-42	1161	! BG106547 602290354F1 NIH_MGC_8

gb_est2:BG5272376	+	48.00	866.50	5.1e-39	761	! BG5272376 602593412F1 NIH_MGC
gb_est2:BI185582	-	43.00	778.41	4.1e-34	374	! BI185582 UNI-P-PN-CY-f-04-0-
gb_est1:BB654969	+	41.00	742.67	4.0e-32	301	! BB654969 BB654969 RIKEN full
gb_est2:BF513673	-	41.00	742.45	4.2e-32	310	! BF513673 UI-H-BW1-amt-h-01-0
gb_est1:BB761195	+	41.00	739.20	6.3e-32	476	! BB761195 BB761195 RIKEN full
gb_est2:BG488694	+	40.00	717.40	1.0e-30	718	! BG488694 602534634F1 NIH_MGC
gb_est1:BB750526	+	39.00	702.43	7.0e-30	439	! BB750526 BB750526 RIKEN full

seq\_name: gb\_est2:BE380437

seq\_documentation\_block:  
LOCUS BE380437 725 bp mRNA linear EST 21-JUL-2000  
DEFINITION 601270510F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3599570 5',  
mRNA sequence.  
ACCESSION BE380437 GI:9325802  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 725)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM8782 row: d column: 03  
High quality sequence stop: 596.  
Location/Qualifiers  
1..725  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_image="3599570"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo df.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 206 a 172 c 179 g 167 t  
ORIGIN

alignment\_scores:  
Quality: 199.00 Length: 199  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-647-377-9 x BE380437

Align seg 1/1 to: BE380437 from: 1 to: 725

146 CysGlyHisProLeuGlnSerArgLysGlyTrpSerGlyProAs 162  
|||||  
1 TGTGGACATCATCCCTCGACAGCTCCCGAAAGGCTGGAGTCTCTGA 50  
162 pVallelleGluLaGlnPheAspSerAspSerGluAspArgHisG 179  
|||||  
51 TGTGATTATAGAGGCTCAGTTTATGACAGCCGACCTCAGAGATG 100  
179 lYasnThrSerGlyLeuValAspGlyValLysLeuSerIleSerThr 195

```
|||||
101 GCAACACCAAGTGGCTGGTTGATGGTGTAAAGAAATTGTCATCTCTACT 150
|||||
196 ProAspArgGlyLysGluAspSerSerThrProValMetLysAspGluAs 212
|||||
151 CTGTGACAGAAAGAGAGATTCTAGTACTCCAGTTATGAAAGATGAGAA 200
|||||
212 nThrProLeuProGlnAspThrArgGlyLeuSerGluLysSerLeuGlnL 229
|||||
201 CACCCCATACACAGACACAAAGAGGCTTATCAGAGAAAGTCACTTCAGA 250
|||||
229 ySerAlaLysValValTyrIleLeuGluLysLysHisSerArgAlaAla 245
|||||
251 AATCAGCAAGAGTGGTTTACATCTTGAGAAAAGCAATCTCAGACAGCA 300
|||||
246 ThrGlyIleLeuLysLeuLeuAlaAspLysAsnSerAspLeuPheLysL 262
|||||
301 ACTGGCATCTCGAAACTCTTTGGCTGATAAGAACAGTGACCTGTTTAA 350
|||||
262 stYrAlaLeuPheSerProSerAspHisArgValProArgIleTyrValP 279
|||||
351 ATACGCCCTGTTTCTCTCTCAGACCAGGAGTACCTAGAAATTTACGTAC 400
|||||
279 roLeuLysAspCysProGlnAspPheMetThrArgProLysAspPheAla 295
|||||
401 CTCTCAAGGACTGTCCCGACGACTTCATGACCCGACCTAAAGACTTTGCC 450
|||||
296 AsnThrLeuPheIleCysArgIleIleAspTrpLysGluAspCysAsnPh 312
|||||
451 AACAGCGTGTTCATCTGCCCATCATAGACTGGAGAGGAGACTGTAAAT 500
|||||
312 eAlaLeuGlyGlnLeuAlaLysSerLeuGlyGlnAlaGlyLeuLeuLup 329
|||||
501 TGCCCTGGGGCACTGGCTAAAGAGTCTTTGGGACGGCTGGTCAATCGAG 550
|||||
329 roGluThrGluGlyIleLeuThrGluTyrGlyValAspPheSerAsp 344
|||||
551 CTGAAACAGAGAGGATACTGACAGATATGGTGTGGACTTCTCTGAT 597
|||||
seq_name: gb_est2:BF450541
```

```
seq_documentation_block:
LOCUS      BF450541
DEFINITION uz67f12.y1 NCI_CGAP Mam6 Mus musculus cDNA clone IMAGE:3674159 5'
            similar to SW:YR86_CAEEL Q09568 HYPOTHETICAL 93.7 KDA PROTEIN
            F48E8.6 IN CHROMOSOME III. [2] SW:YR86_CAEEL ; mRNA sequence.
ACCESSION  BF450541
VERSION    BF450541.1 GI:11516710
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 681)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT   Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: Washington University Genome Sequencing Center
            found through the I.M.A.G.E. Consortium/LLNL at:
            Image.llnl.gov/image/html/iresources.shtml
```

```
MGI:1434927
Seq primer: -40RP from Gibco
High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source            1..681
```

```
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3674159"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT  160 a 206 c 169 g 144 t 2 others
ORIGIN
```

```
alignment_scores:
    Quality: 195.00      Length: 195
    Ratio: 1.000        Gaps: 0
    Percent similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-09-647-377-9 x BF450541
Align seg 1/1 to: BF450541 from: 1 to: 681
```

```
489 LysLeuSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLys 505
|||||
1 AAACAGCTAGCAGCAGCATGCCAGAGCATGATCGAAATCCAACTGAGAA 50
|||||
505 stleProGluGluGluLeuProProIleSerProGluHisSerValGluG 522
|||||
51 GATCCCTGAGGAGAGCTTCCCCCAATTTCTCCAGAGCACAGCGTCGAGG 100
|||||
522 luValHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArg 538
|||||
101 AGGTGACCAGGCAGTCTCGAACCTGCACAGCATTGCAAAAGCAACTCCGC 150
|||||
539 ArgGlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAl 555
|||||
151 CGCCACGCGCTTTAGATAGCGGCACACTCCGTTTAGATCAGCTGAAGCTGC 200
|||||
555 apheThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrG 572
|||||
201 TTTTACTCTGGACCATGAGACTGGATTGGCTCAAGGATGTCACATCTATG 250
|||||
572 luTyrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsn 588
|||||
251 AGTACGAGACAGCAACAAGCTTTAGAGAGATTCATGCTCTCTGGCCAAC 300
|||||
589 MetAlaValAlaHisLysLysIlePheArgThrPheProGluGlnAlaLeu 605
|||||
301 ATGGGGGTGGGCCCAAGATCTTCCGCACCTTCCCTGAGCAGGCCCTGCT 350
|||||
605 uArgArgHisProProGlnThrLysMetLeuSerAspLeuValGluP 622
|||||
351 GCGCCGGCATCCCCACCACAGAGATGCTCAGTGACCTGGTGAGT 400
|||||
622 heCysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeu 638
|||||
401 TCTGTGACCACTGGGGCTGCCCATGATGTCAGCTCTGACAGGGGCCCTA 450
|||||
639 AsnLysSerLeuThrLysThrPheGlyAspAspLysTyrSerLeuAlaAr 655
|||||
451 AATAAAAGCCTGACTAAGACATTTGGAGATGACAAGTACTCTCTGCCCCG 500
|||||
655 gLysGluValLeuThrAsnMetTyrSerArgProMetGlnMetAlaLeuT 672
|||||
501 GAAGAGGTGCTCACCACATGTACTCCCGGCCCATGCAGATGGCAGTGT 550
|||||
672 yrPheCysSerGlyMetLeuGlnAspGlnGluGln 683
|||||
```

```
551 ACTTCTGCTGTGGATGCTGCAGGACCAAGAGCAG 585
seq_name: gb_est2:BI988256
seq_documentation_block:
LOCUS BI988256 600 bp mRNA linear EST 20-DEC-2001
DEFINITION 3223-16 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
mRNA sequence.
ACCESSION BI988256
VERSION BI988256.1 GI:17959235
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 600)
M,X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
Nucleic Acids Res. 29 (24), 4983-4993 (2001)
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
FEATURES
Location/Qualifiers
source
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
BASE COUNT 143 a 171 c 149 g 137 t
ORIGIN
alignment_scores:
Quality: 186.00 Length: 200
Ratio: 0.935 Gaps: 1
Percent Similarity: 99.500 Percent Identity: 99.500
alignment_block:
US-09-647-377-9 x BI988256 ..
Align seq 1/1 to: BI988256 from: 1 to: 600
401 ThrPheGluValGlyValHisIleAlaAspValSerTyrPheValProG1 417
|||||
2 ACCTTCGAAGTGGCGTCCACATCGCGATGTGAGTTACTTTGTTCCTGA 51
|||||
417 uGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrL 434
|||||
52 GGGATCCTCTTTGGATAAGTAGCTGCTGAGAGAGCCACAAAGTGCTACT 101
|||||
434 euValGlnLysValValProMetLeuProArgLeuLeuCysGluGluLeu 450
|||||
102 TGGTCCAGAGAGGTGGTCCCATGCTTCCAGGCTTCTGTGTGAGGAATC 151
|||||
451 CysSerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLy 467
|||||
152 TGCAGGCTCAACCCATGACTGACAAAGCTGACCTTCTGTGATCGGAA 201
|||||
467 sLeuThrProGluGlyLysIleLeuGluGluTyrPheGlyArgThrIleI 484
|||||
202 GCTGACCCCTGAAGGCAAGATCCTTGAAGAGTGGTTTGGCCGACATCA 251
|||||
484 leArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGlu 500
|||||
252 TCCGTTCTTGCACCAACTGAGCTACGACCATGCCAGAGCATGATCGAA 301
|||||
501 AsnProThrGluLysIleProGluGluLeuProIleSerProG1 517
|||||
302 AATCCAACTGAGAAGATCCCTGAGGAAGAGCTTCCCCCAATTCTCCAGA 351
|||||
517 uHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleA 534
|||||
352 GCACAGCTCGAGGAGGTGCACCGAGCTCTGAACCTGCACAGCATTG 401
|||||
534 laLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAsp 550
|||||
402 CAAAGCAACTCCGCCGCCAGGCTTTGTAGATGGCCACTCCGTTTAGAT 451
|||||
551 GlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnG1 567
|||||
452 CAGCTGAAGCTTCTTTTACTCTGGACCATGAGACTGGATTGCTCAAAG 501
|||||
567 yCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPhe 583
|||||
502 ATGTCACTATCTAGTACCCGAGACCAACAAGCTTGTAGAGGAGTTC 551
|||||
584 MetLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhe 599
|||||
552 ATGCTCTGCGCAACATGCGGTGGCCACCAAGATCTTCGCACCTTC 599
|||||
seq_name: gb_est2:BI661242
seq_documentation_block:
LOCUS BI661242 682 bp mRNA linear EST 12-SEP-2001
DEFINITION 603304587F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350461 5',
mRNA sequence.
ACCESSION BI661242
VERSION BI661242.1 GI:15575478
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 682)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1890 row: m column: 22
High quality sequence stop: 682.
Location/Qualifiers
1..682
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5350461"
/clone_lib="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert 2.5 Kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furth, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
Library."
BASE COUNT 195 a 140 c 187 g 159 t 1 others
ORIGIN
alignment_scores:
```

Quality: 178.00 Length: 178  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x BI661242 ..

Align seg 1/1 to: BI661242 from: 1 to: 682

```
5 AsptYrLysLeuAsnLeuArgSerProGlyThrProArgGlyValSerSe 21
|||||
106 GACTACAAAGCTGAACCTTCCTCGGGACCCACAGAGTGTCCTC 155
|||||
21 rValValGlyProSerAlaValGlyAlaSerProGlyAspLysSerL 38
|||||
156 TGTGTTGGCCGAGTGCTGTTGGTCTTCGCCAGGTGACAAAAGTCAA 205
|||||
38 ysAsnLysSerMetArgGlyLysLysSerIlePheGluThrTyrMet 54
|||||
206 AGAACAAAGTCCATCGGAGGGAAGAAAAGAGCATATTTGAAACCTACATG 255
|||||
55 SerLysGluAspValSerGluGlyLeuLysArgGlyThrLeuIleGlnI 71
|||||
256 TCCAGAGGAGGAGTTCRGAAGGCTTGAAGAGGAACTTATPCACGGG 305
|||||
71 yValLeuArgIleAsnProLysLysPheHisGluAlaPheIleProSerP 88
|||||
306 TGTATTGAGAATCAACCAAGAAGTTTCATGAAGCCTTCATTCCTTC 355
|||||
88 rAspGlyAspArgAspIlePheIleAspGlyValValAlaArgAsnArg 104
|||||
356 CGAGTGATCGGGACATTTTATGTGAGGAGTTGTCGCTGTAATAGA 405
|||||
105 AlaLeuAsnGlyAspLeuValValLysLysLeuLeuProGluAspGlnTr 121
|||||
406 GCCTTAAATGGGACCTTGTGTGTGTAAACTGCTTCCTGAGGATCAGTG 455
|||||
121 pLysAlaValLysProGluSerAsnAspLysGluIleGluAlaThrTyrG 138
|||||
456 GAAGCGAGTTAAACCGAGAGCANTGCANAGAATAGAGCTACTTATG 505
|||||
138 luAlaAspIleProGluGluGlyCysGlyHisHisProLeuGlnGlnSer 154
|||||
506 AAGCTGACATCCCTGAAGAGGGCTGTGGACATCACCCCTCGACAGTCC 555
|||||
155 ArgLysGlyTrpSerGlyProAspValIleIleGluAlaGlnPheAspAs 171
|||||
556 CGGAAGGCTGGAGTGCTCTCATGTATTATAGAGCTCGTTTGATGA 605
|||||
171 pSerAspSerGluAspArgHisGlyAsnThrSer 182
|||||
606 CAGCGACTCAGAAGATAGACATGGCAACACCAAGT 639
|||||
```

seq\_name: gb\_est2:BI078196

seq\_documentation\_block:

LOCUS BI078196 695 bp mRNA linear EST 20-JUN-2001  
DEFINITION 602672520r1 NCI\_CGAP\_Mam2 Mus musculus cdna clone IMAGE:5004341.5',  
.. mRNA sequence.

ACCESSION BI078196

VERSION BI078196.1 GI:14496526

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 695)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM11042 row: h column: 06

High quality sequence stop: 689.

#### FEATURES

source

Location/Qualifiers

1..695

/organism="Mus musculus"

/strain="FVB/N-3"

/db\_xref="taxon:10090"

/clone="IMAGE:5004341"

/clone\_lib="NCI\_CGAP\_Mam2"

/tissue\_type="tumor, biopsy sample"

/dev\_stage="5 months"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT 191 a 151 c 200 g 153 t

ORIGIN

alignment\_scores:

Quality: 175.00

Ratio: 0.931

Percent similarity: 99.471

Percent Identity: 99.471

Gaps: 1

Length: 189

US-09-647-377-9 x BI078196 ..

Align seg 1/1 to: BI078196 from: 1 to: 695

```
1 MetAsnHisProAspTyrLysLeuAsnLeuArgSerProGlyThrProAr 17
|||||
126 ATGAACCATCTCTGACTACAGCTGAACCTTCGCTCTCCGGGGACCCCGAG 175
|||||
17 gGlyValSerSerValValGlyProSerAlaValGlyAlaSerProGlyA 34
|||||
176 AGGTGTGCTCTGTGGTTGGCCGAGTCTGTGTGCTTCGCCAGGTG 225
|||||
34 spLysLysSerLysAsnLysSerMetArgGlyLysLysLysSerIlePhe 50
|||||
226 ACAAAAAGTCAAGAACCAAGTCCATGCGAGGGAAGAAAAAGAGCATATT 275
|||||
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67
|||||
276 GAAACCTACATGTCGAAGGAGGATGTTTCAGAAGGCTTGAAGAGGGAAC 325
|||||
67 rLeuIleGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84
|||||
326 ACTTATCCAGGGTGATTTGAGAATCAACCAAGAGAGTTTCATGAAGCCT 375
|||||
84 heileProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100
|||||
376 TCATTCCTTCCTCGGATGTGATCGGGACATTTTATTGATGGAGTTGTT 425
|||||
101 AlaArg.AsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuP 117
|||||
426 GCTCGTAAATAGAGCCCTTAATGGGACCTTGTGGTGTGTAAGACTGCTTC 475
|||||
117 roGluAspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIle 133
|||||
476 CTGAGGATCAGTGGAGGCGAGTTAAACCCAGAGAGCAATGACAAAGAAATA 525
|||||
134 GluAlaThrTyrGluAlaAspIleProGluGluGlyCysGlyHisHisPr 150
|||||
526 GAAGCTACTTATGAAGCTGACATCCCTGGAAGAGGGCTGTGGACATCACC 575
|||||
150 oLeuGlnGlnSerArgLysGlyTrpSerGlyProAspValIleIleGluA 167
```



Average insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999). Note: this is a NCI\_CGAP Library."

BASE COUNT  
ORIGIN

351 a 259 c 326 g 221 t 1 others

alignment\_scores:  
Quality: 158.00 Length: 158  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x BE650394 ..

Align seg 1/1 to: BE650394 from: 1 to: 1158

```
1 MetAsnHisProAspTyrLysLeuAsnLeuArgSerProGlyThrProAr 17
|||||
131 ATGAACCATCTCTGACTCAAGCTCAACCTTCGCTCTCGGGGACCCGAG 180
|||||
17 gGlyValSerSerValValGlyProSerAlaValGlyAlaSerProGlyA 34
|||||
181 AGGTGTCTCTCTGTTGGCCGAGTCTCTGTGTGCTTCGCCAGGTG 230
|||||
34 splLysLysSerLysAsnLysSerMetArgGlyLysLysSerIlePhe 50
|||||
231 ACAAAAGTCAAAAGAACCAAGTCCATCGAGGGAAGAAAAGAGCATATT 280
|||||
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67
|||||
281 GAAACCTACATGTCGAAGGAGGATGTTTCAGAGGCTTGAAGAGAGAAC 330
|||||
67 rLeuLeuGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84
|||||
331 ACTTATCCAGGGTGTATTGAGATCAACCAAGAAAGTTTCATGAAGCCT 380
|||||
84 heIleProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100
|||||
381 TCATTCTCTCTCCGGATGGTGATCGGACATTTTATTGATGGAGTTGTT 430
|||||
101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuPr 117
|||||
431 GCTCGTAATAGAGCCCTTAAATGGGACCTTGTGGTTGTAAACCTGCTTC 480
|||||
117 oGluAspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIleG 134
|||||
481 TCAGGATCAGTGGGAAGCGACTTAAACACAGAGAGCAATGACAAAGAAATAG 530
|||||
134 luAlaThrTyrGluAlaAspIleProGluGluGlyCysGlyHisHisPro 150
|||||
531 AAGCTACTTATGAGCTGACATCCCTGGAAGAGGGCTGTGGACATCACCCC 580
|||||
151 LeuGlnGlnSerArgLysGlyTrp 158
|||||
581 CTGACGAGCTCCCGAAGAGGCTGG 604
```

seq\_name: gb\_est2:BE380350

seq\_documentation\_block:

LOCUS BE380350 444 bp mRNA linear EST 21-JUL-2000  
DEFINITION 601270414F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3599763 5',  
mRNA sequence.

ACCESSION BE380350

VERSION BE380350.1 GI:9325715

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 444)

AUTHORS

TITLE NIH-MGC <http://mgc.nci.nih.gov/>.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8782 row: 1 column: 04  
High quality sequence stop: 442.

Location/Qualifiers

source

1. .444

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:3599763"

/clone\_lib="NCI\_CGAP\_Mam1"

/tissue\_type="tumor, biopsy sample"

/dev\_stage="10 months, virgin"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 124 a 108 c 105 g 107 t

ORIGIN

alignment\_scores:

Quality: 147.00 Length: 147

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x BE380350 ..

Align seg 1/1 to: BE380350 from: 1 to: 444

```
227 LeuGlnLysSerAlaLysValValTyrIleLeuGluLysLysHisSerAr 243
|||||
2 CTTCAGAAATACAGCAAGGTGTTTACATCTCGAGAAAAAGCATTCGCG 51
|||||
243 gAlaAlaThrGlyIleLeuLysLeuLeuAlaAspLysAsnSerAspLeup 260
|||||
52 AGCAGCAACTGGCATCTCTGAAACTCTTGGCTGATAGACACAGTGCCTGT 101
|||||
260 heLysLysTyrAlaLeuPheSerProSerAspHisArgValProArgIle 276
|||||
102 TTAAGAAATACGCCCTGTTTCTTCCTTCAGACCACCGAGTACCTAGATT 151
|||||
277 TyrValProLeuLysAspCysProGlnAspPheMetThrArgProLysAs 293
|||||
152 TAGCTACCTCTCAGGACTGTCCCGAGGACTTCATGACCCGACCTAAGA 201
|||||
293 pPheAlaAsnThrLeuPheIleCysArgIleIleAspTrpLysGluAspC 310
|||||
202 CTTTGCACACCGCTGTTTCATCTGCCGCATCATAGCTGGAGGAGGACT 251
|||||
310 ysAsnPheAlaLeuGlyClnLeuAlaLysSerLeuGlyClnAlaGlyGlu 326
|||||
252 GTAATTTGCCCTGGGGCACTGGCTAAAGAGTCTTTGGCAGGCTGGTGA 301
|||||
327 IleGluProGluThrGluGlyIleLeuThrGluTyrGlyValAspPheSe 343
|||||
302 ATCGAGCCTGAAACAGAGGAGTACTGACAGATATGTTGGACTTCTC 351
|||||
343 raspPheSerSerGluValLeuGluCysLeuProGlnSerLeuProTrpT 360
|||||
352 TGATTTCTCTTCAGAAAGTTCTTGAATGATCTCTCCCTCAAAGCCTGCC 401
```



samples: Lothar Hennighausen/Priscilla Furth, NIH  
Reference for transgenic model: Li et al., Cell Growth and  
Differentiation 7, 3-11 (1996). Note: this is a NCI\_CGAP  
Library."

BASE COUNT 171 a 194 c 174 g 162 t  
ORIGIN

alignment\_scores:  
Quality: 126.00 Length: 126  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-647-377-9 x BG864079 ..  
Align seg 1/1 to: BG864079 from: 1 to: 701

```

328 GluProGluThrGluGlyIleLeuThrGluTyrGlyValAspPheSerAs 344
|||||
2 GAGCCTGAAACAGAGAGGATCTGACAGAAATATGTTGGACTTCTCTGA 51
|||||
344 pPheSerSerGluValLeuGluCysLeuProGlnSerLeuProTrpThrI 361
|||||
52 TTTCTCTTCAGAAATCTTGAATGTCTCCCTCAAAAGCCTGCCCTGGACAA 101
|||||
361 leProProAspGluValGlyIleArgAspLeuArgLysAspCysIle 377
|||||
102 TCCCACTGATGAGGTGGGACAGAGAGACCTAAGGAAAGACTGTATC 151
|||||
378 PheThrIleAspProSerThrAlaArgAspLeuAspAspAlaLeuAlaCy 394
|||||
152 TTCACCATTTGATCATCAACTGTCGCGACCTTGATGATGCCCTCGCCTG 201
|||||
394 sArgArgLeuThrAspGlyThrPheGluValGlyValHisIleAlaAspV 411
|||||
202 CAGCGCGCTCACTGATGGACCTTCGAAATGGGCGTCCACATGCCCATG 251
|||||
411 alSerTyrPheValProGluGlySerSerLeuAspLysValAlaAlaGlu 427
|||||
252 TGAGTTACTTTGTTCTCGAGGGATCTCTTTGGATAAAGTAGCTGTGAG 301
|||||
428 ArgAlaThrSerValTyrLeuValGlnLysValValProMetLeuProAr 444
|||||
302 AGAGCCACAAAGTGTCTACTTGTGTCACAGAGGTGGTCCCATGCTTCCAG 351
|||||
444 GluLeuCysGluGluLeuCysSerLeu 453
|||||
352 GCTTCTGTGTGAGGAACCTCTGCAGCCTC 379

```

seq\_name: gb\_est2:BE910893

seq\_documentation\_block:  
LOCUS BE910893 506 bp mRNA linear EST 29-SEP-2000  
DEFINITION 601661970F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3962249 5',  
mRNA sequence.  
ACCESSION BE910893  
VERSION BE910893.1 GI:10407877  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS 1 (bases 1 to 506)  
TITLE NIH-MGC http://mgc.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@email.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA library Preparation: Life Technologies, Inc.  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9128 row: k column: 18  
High quality sequence stop: 506.

FEATURES  
Location/Qualifiers

1..506  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3962249"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, Nih"

BASE COUNT 136 a 105 c 145 g 120 t  
ORIGIN

alignment\_scores:  
Quality: 122.00 Length: 122  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-647-377-9 x BE910893 ..  
Align seg 1/1 to: BE910893 from: 1 to: 506

```

1 MetAsnHisProAspTyrIleLeuAsnLeuArgSerProGlyThrProAr 17
|||||
105 ATGAACCATCTCTGACTACAACTGAACCTTCGGTCTCCGGGGAGCCCCAG 154
|||||
17 gGlyValSerSerValValGlyProSerAlaValGlyAlaSerProGlyA 34
|||||
155 AGGTGTGTCTCTGTGTGGTGGCCCGAGTGTGTGGTGTTCGCCAGGTG 204
|||||
34 sPlyLysSerLysAsnLysSerMetArgGlyLysLysLysSerIlePhe 50
|||||
205 ACAAAGTCAAGAACCAAGTCCATGCGGGGAAGAAAGAGCATATTT 254
|||||
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67
|||||
255 GAAACCTACATGTCCAAGGAGGATGTTTCAGAGGCTTGAAGAGGAGAC 304
|||||
67 rLeuIleGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84
|||||
305 ACTATCCAGGGTGTATTGAGAATCAACCCAAAGAGTTTCATGAAGCCT 354
|||||
84 heileProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100
|||||
355 TCATTCCTCTCCGGATGTCGATCGGACATTTTATTGATGGAGTGT 404
|||||
101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuPr 117
|||||
405 GCTCGTAATAGAGCCTTAATGGGGACCTGTGTGTGTGTAAACCTGCTTC 454
|||||
117 oGluAspGlnTrpLys 122
|||||
455 TGAGGATCAGTGAAG 470

```

seq\_name: gb\_est1:BB636311

seq\_documentation\_block:  
LOCUS BB636311 684 bp mRNA linear EST 31-AUG-2001  
DEFINITION BB636311 RIKEN full-length enriched, 0 day neonate thymus Mus  
musculus cDNA clone A430109E24 5', mRNA sequence.  
ACCESSION BB636311  
VERSION BB636311.1 GI:15400918





/clone\_lib="Stratagene mouse heart (#937316)"

/sex="pooled"

/tissue\_type="heart"

/dev\_stage="13 day embryos"

/lab\_host="SOLR (kanamycin resistant)"

/note="Organ: heart; Vector: pBluescript SK-; Site 1:

EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:

Oligo dn: 93 pooled NIH/Swiss 13 day embryo hearts.

Average insert size: 1.0 Kb; Uni-ZAP XR Vector; -5'

adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor

sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT 123 a 95 c 126 g 107 t

ORIGIN

alignment\_scores:  
Quality: 115.00 Length: 115  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x AI597514 ..

Align seg 1/1 to: AI597514 from: 1 to: 451

```

1 MetAsnHisProAspTyrLysLeuAsnLeuArgSerProGlyThrProAr 17
|||||
105 ATGAACCATTCCTGACTACAGCTGAACCTTCGGTCTCGGGGACCCCGAG 154
|||||
17 gGlyValSerSerValAlGlyProSerAlaValGlyAlaSerProGlyA 34
|||||
155 AGGTGTGCTCTGTGTGTGGCCCGAGTGTGTGTGTGTGTGTGTGTGTGT 204
|||||
34 spLysLysSerLysLysSerMetArgGlyLysLysLysSerIlePhe 50
|||||
205 ACAAAAAGTCNAAGAACAAAGTCATGCGAGGGAAGAAAGAGCATATTT 254
|||||
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67
|||||
255 GAACCTACATGTCACAGGAGGATGTTTCAGAGGCTTCAAGAGAGGAAC 304
|||||
67 rLeuIleGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84
|||||
305 ACTATCCAGGGTGATTGAGAATCAACCCAAAGAGTTTCATGAAGCCT 354
|||||
84 helleProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100
|||||
355 TCATTCTCTCCGGATGTCGATCGGACATTTTATTGATGGAGTTGTT 404
|||||
101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValValLysLeu 115
|||||
405 GCTCGTAATAGACCTTAATAGGGGACCTTGTTGTTGTAACACTG 449
|||||

```

seq\_name: gb\_est2:BG804017

seq\_documentation\_block:

LOCUS BG804017 600 bp mRNA linear EST 20-DEC-2001

DEFINITION 0244-55 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA.

FEATURES

ACCESSION BG804017

VERSION BG804017.1 GI:17950931

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 600)

White, R.A., Beremand, P.D., Thomas, I.D., Gan, L. and Klein, W.H.

Gene expression in the developing mouse retina by EST sequencing

and microarray analysis

Nucleic Acids Res. 29 (24), 4983-4993 (2001)

CONTACT: Klein WH

Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 792 0329.

FEATURES

source

Location/Qualifiers

1..600

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

/tissue\_type="neural retina"

/dev\_stage="embryonic day 14.5 post-fertilization"

BASE COUNT 147 a 173 c 151 g 126 t

ORIGIN

alignment\_scores:  
Quality: 109.00 Length: 109  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x BG804017 ..

Align seg 1/1 to: BG804017 from: 1 to: 600

```

490 LeuSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysIrl 506
|||||
129 CTGAGCTACGACCATGCCAGAGCATGATCGAAATCAACTGAGAAGAT 178
|||||
506 eProGluGluGluLeuProProIleSerProGluHisSerValGluGluV 523
|||||
179 CCTCTGAGGAGAGGCTTCCCCCAATTTCTCCAGAGCACAGCGTCGAGGAG 228
|||||
523 aHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArg 539
|||||
229 TGCACCGAGGAGTCTCTGACCTGCACAGCATTCGAAAGCAACTCCGCCGC 278
|||||
540 GlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaph 556
|||||
279 CAGCGCTTTTGTAGATGGCGCACTCCGTTTAGATCAGCTGAAGCTTGCTTT 328
|||||
556 eThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrGluT 573
|||||
329 TACTCTGGACCATGAGACTGGATTCCTCAGGATGTCACATCTATGAGT 378
|||||
573 YrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsnMet 589
|||||
379 ACCGAGACAGCAACAAGCTTGTAGAGGAGTTTCATGCTCTGGCCACATG 428
|||||
590 AlaValAlaHisLysLysIlePheArgThr 598
|||||
429 GCGGTGGCCCAAGATCTTCCGCACC 455

```



OM of: US-09-647-377-9 to: EST:\* out\_format : pfs

Date: Jul 8, 2002 12:04 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed-p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09647377/runat\_08072002\_075944\_22277/app\_query.fasta\_l.940  
-DB=EST -Qfmt=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-LIST=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi  
-LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09647377 -CGNL\_l\_1.7845  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-647-377-9

Query length: 870

Database: EST.\*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1635.870000

score\_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_est2:BE791331	+	1193.50	2196.17	4.4e-113	887 ! BE791331 601582710F1 NIH_MGC_7
gb_est2:BG825206	+	1162.00	2139.67	6.2e-110	735 ! BG825206 602747617F1 NIH_MGC_17
gb_est2:BF450541	+	1159.00	2134.96	1.1e-109	681 ! BF450541 u267f12.y1 NCI_CGAP_M
gb_est2:BG864079	+	1158.00	2132.77	1.5e-109	701 ! BG864079 602797865F1 NIH_CGAP_M
gb_est2:BG746662	+	1153.00	2123.04	5.2e-109	727 ! BG746662 602703970F1 NIH_MGC_15
gb_est2:BM012844	+	1114.50	2048.12	7.8e-105	961 ! BM012844 603637574F1 NIH_MGC_47
gb_est2:BE380437	+	1076.00	1979.67	5.1e-101	725 ! BE380437 601270510F1 NCI_CGAP_M
gb_est2:BG106547	+	1072.00	1966.80	2.6e-100	1161 ! BG106547 602290354F1 NIH_MGC_8
gb_est2:BG750784	+	1050.50	1932.40	2.2e-98	711 ! BG750784 602706985F1 NIH_MGC_43
gb_est2:BF182212	+	1037.00	1904.57	7.7e-97	898 ! BF182212 601804759F1 NCI_CGAP_M
gb_est2:BF114256	+	1021.00	1875.75	3.1e-95	825 ! BF114256 602862446F1 NIH_MGC_17
gb_est2:BF1988256	+	1019.00	1875.70	3.1e-95	600 ! BF1988256 3223-16 Mouse E14.5 re
gb_est2:BG804017	+	1015.00	1868.25	8.1e-95	600 ! BG804017 0244-55 Mouse E14.5 re
gb_est2:BE284271	+	1015.00	1867.83	8.6e-95	622 ! BE284271 601087339F1 NCI_CGAP_M
gb_est2:BF1650394	+	1002.50	1837.40	4.3e-93	1158 ! BF1650394 603296125F1 NIH_CGAP_M
gb_est2:BF614228	+	986.00	1811.73	1.1e-91	746 ! BF614228 de03f05.y1 Wellcome CR
gb_est2:BF182212	+	982.00	1805.31	2.6e-91	682 ! BF182212 603304587F1 NIH_CGAP_M
gb_est2:BF1078196	+	979.00	1799.51	5.5e-91	695 ! BF1078196 602872520F1 NCI_CGAP_M
gb_est2:BE545087	+	971.50	1785.51	3.3e-90	697 ! BE545087 601077821F1 NIH_MGC_12
gb_est1:BE131806	+	926.00	1701.23	1.6e-85	670 ! BE131806 db38c11.y1 Blackshear
gb_est1:AL652258	+	918.00	1685.78	1.2e-84	703 ! AL652258 AL652258 XGC-gastrula
gb_est1:BF304859	+	914.00	1675.65	4.4e-84	887 ! BF304859 60188211F1 NIH_MGC_17
gb_est2:BE534295	+	894.00	1644.53	2.4e-82	521 ! BE534295 601235354F1 NCI_CGAP_M
gb_est2:BF684505	+	868.00	1595.18	1.3e-79	565 ! BF684505 185915 MARC 4BOV Bos t
gb_est2:BF198482	+	867.00	1592.85	1.8e-79	588 ! BF198482 248425 MARC 2PTG Sus s
gb_est2:BF204262	+	827.50	1517.40	2.8e-75	693 ! BF204262 601867724F1 NIH_MGC_17
gb_est2:BF608541	+	823.00	1512.00	5.7e-75	535 ! BF608541 307159 MARC 1PTG Sus s
gb_est1:AJ393871	+	815.50	1493.68	6.0e-74	781 ! AJ393871 AJ393871 rikfz426 Gallu
gb_est1:BM615143	+	781.00	1430.82	1.9e-70	692 ! BM615143 BB615143 RIKEN full-le
gb_est2:BF548574	+	778.00	1429.86	2.8e-70	761 ! BF548574 UI-R-A0-at-d-12-0-UI-R
gb_est2:BF548574	+	778.00	1429.37	2.3e-70	483 ! BF548574 UI-R-A0-at-d-12-0-UI-R
gb_est2:BF045061	+	775.00	1428.70	2.5e-70	512 ! BF045061 BP250017820F1 Soares
gb_est2:BF380350	+	775.00	1424.76	4.1e-70	444 ! BF380350 601270414F1 NCI_CGAP_M
gb_est1:AW641271	+	763.00	1399.71	1.0e-68	561 ! AW641271 cm05d06.y1 Blackshear
gb_est2:BG877875	+	757.00	1390.08	3.5e-68	491 ! BG877875 QV4-LT0016-240200-110
gb_gss:GNS04MLJ	-	737.00	1345.41	1.1e-65	935 ! AL297588 Tetradon nigriviridis
gb_est1:AA756851	+	728.00	1336.00	3.6e-65	494 ! AA756851 vr20c02.r1 Barstead md
gb_est2:BG878140	+	713.00	1305.77	1.7e-63	603 ! BG878140 MARC-PT0052-150200-101
gb_est2:BM490835	+	709.00	1300.47	3.4e-63	500 ! BM490835 pfp2n.pk004.j4 Normal
gb_est1:BB620465	+	705.50	1291.06	1.1e-62	643 ! BB620465 BB620465 RIKEN full-le

gb\_est2:BF047956 + 700.00 1283.58 3.0e-62 506 ! BF047956 dc85g10.y1 NICHHD XG  
gb\_est1:AL662277 + 700.00 1283.13 3.2e-62 526 ! AL662277 AL662277 XGC-neurul  
gb\_est2:BE886139 + 698.50 1275.28 8.7e-62 816 ! BE886139 60150925F1 NIH\_MGC  
gb\_est1:BB636311 + 670.00 1224.24 6.1e-59 684 ! BB636311 BB636311 RIKEN full  
gb\_est2:BI848955 + 663.50 1216.09 1.7e-58 485 ! BI848955 471556 MARC 2BOV Bo

seq\_name: gb\_est2:BE791331

seq\_documentation\_block:

LOCUS BE791331 887 bp mRNA linear EST 20-SEP-2000  
DEFINITION NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3937083 5',  
mRNA sequence.  
ACCESSION BE791331  
VERSION BE791331.1 GI:10212529  
KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 887)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue procurement: DCTD/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCW782 row: c column: 04

High quality sequence stop: 750.

Location/Qualifiers

1..887

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3937083"

/clone\_lib="NIH\_MGC\_7"

/tissue\_type="small cell carcinoma"

/cell\_line="MGC3"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pOTB7; Site: 1; XhoI; Site: 2;

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GCGACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 238 a 205 c 229 g 215 t

ORIGIN

alignment\_scores:

Quality: 1193.50

Ratio: 4.356

Percent Similarity: 94.483

Percent Identity: 83.448

alignment\_block:

US-09-647-377-9 x BE791331

Align seg 1/1 to: BE791331 from: 1 to: 887

197 AspArgGlyLysGluAspSerSerThrProValMetLysAspGluAnThr 213

11 GAGAAAGGAGAGAGATGTGTATGATGACCCGGTTACAAAAGATGAGACCAC 60

213 rProtleProGlnAspThrArgGlyLeuSerGluLysSerLeuGlnLys 230

61 CPGCATTTCAACAGACACAGAGCTTTATCGGAGAAATCCCTCGCAAGAT 110

230 exAlaLysValItyrIleLeuGluLysLysHisSerArgAlaAlaThr 246

```

111 CAGCAAGAGTGGTTTACATCTTGAGAAAAAACAATTCGAGCAGCAACC 160
247 GlyIleLeuLysLeuLeuAlaAspLysAsnSerAspLeuPheLysLysTy 263
|||||
161 GCGTTCCTCAAACTCTTGGCTGATAGAACACAGCAACTGTTTAGGAANA 210
263 rAlaLeuPheSerProSerAspHisArgValProArgIleTyrValProL 280
211 CGCCCTGTTTCTCCCTCAGACACCCGAGTGCCTAGAATTTATGTGCCTC 260
280 euLysAspCysProGlnAspPheMetThrArgProLysAspPheAlaAsn 296
261 TCAAGGAGCTGCCCGCAGCACTTTGGCCAGCCCTTAAAGATTATGCCAAC 310
297 ThrLeuPheLleCysArgIleLeuAspTrpLysGluAspCysAsnPheAl 313
311 ACACGTGTTTCATCTGCCGCATTTGGACTGGAAGGAGGACTGCAATTTTGC 360
313 aLeuGlyGlnLeuAlaLysSerLeuGlyGlnAlaGlyGluIleGluProG 330
361 CCGTGGGCGAGCTGGCTAAGAGCTCTGGCGAGCTGGTGAATTGAGCCTG 410
330 luThrGluGlyIleLeuThrGluTyrGlyValAspPheSerAspPheSer 346
411 AAACAGAGAAGGAATACTAACAGAGTATGGCGTGGATTTCTGTGATTTCT 460
347 SerGluValLeuGluCysLeuProGlnSerLeuLeuProThrPheProPr 363
461 TCAGAGTCTAGNAATGCTTCTCCACGCCCTGCATGGACATTCACCC 510
363 aspGluValGlyLysArgAspLeuArgLysAspCysIlePheThrI 380
511 AGAGAGGATTCACAGCAAGAGAGGATTTAAGAAAGAGCTGTATCTCACCA 560
380 LeAspProSerThrAlaArgAspLeuAspAspAlaLeuAlaCysArgArg 396
561 TTGACCCATCAACGCCGCGAGACCTCGATGATGCCCTCTCTGCAAGCCA 610
397 LeuThrAspGlyThrPheGluValGlyValHisIleAlaAspValSerTy 413
611 CTCGCTACCGGCACTTCAAGTGGGAGTTCCATTCGTGCGTGGATTA 660
413 rPheValProGluGlySerSerLeuAspLysValAlaAlaGluArgAla 430
661 CTTTGTTCGGAGGAGTCTGATCTGGATAAAGTGGTCCCGAAGAGGCTA 710
430 hr.SerValTyrLeuValGlnLysValValProMetLeuProArgLeuLe 446
711 CAAGAGCTCTACTTGTGTTCAAGGTGGTCCCATGCTTCCAGGCTGTG 759
446 uCysGluGluLeuCysSerLeuAsnProMetThrAspLysLeuThrPheS 463
760 .TGTGAAGAGCTGTGCAGGCTTAACCCAACTGTCGAAAGAGCTGACCTTC 808
463 erValIleTrpLysLeuThrProGluGlyLysIleLeuGluGluTrpPhe 479
809 T.GTGATTTGGGCACCTGGC.CCAGAGGCAAAAACCTTGAAAGG...GTT 854
480 GlyArgThrIleIleArg 485
855 GGGCGGACCATATCCGT 872

seq_name: gb_est2:BG825206

seq_documentation_block:
LOCUS BG825206 735 bp mRNA linear EST 22-MAY-2001
DEFINITION 602747617P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4900492 5',
mRNA sequence.
ACCESSION BG825206
VERSION BG825206.1 GI:14172793
KEYWORDS EST.
SOURCE human.

```

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1. (bases 1 to 735)

## AUTHORS

NIH-MGC <http://mgc.ncl.nih.gov/>.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: AFCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LICM1794 row: a column: 05

High quality sequence stop: 732.

## FEATURES

Location/Qualifiers

1..735

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4900492"

/clone\_lib="NIH\_MGC\_17"

/tissue\_type="rhabdomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;

Site\_2: XhoI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAGAG(G). Size-selected &gt;500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

## BASE COUNT

167 a 217 c 200 g 151 t

## ORIGIN

## alignment\_scores:

Quality: 1162.00 Length: 244

Ratio: 4.903 Gaps: 0

Percent Similarity: 97.131 Percent Identity: 90.574

## alignment\_block:

US-09-647-377-9 x BG825206 ..

Align seg 1/1 to: BG825206 from: 1 to: 735

394 CysArgArgLeuThrAspGlyThrPheGluValGlyValHisIleAlaAs 410

||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

3 TGCAGCCACTCGCTGCGAGCACTTCAAGTGGAGTTCACATGCTGA 52

410 pValSerTyrPheValProGluGlySerSerLeuAspLysValAlaAlaG 427

||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

53 CDTGAGTTACTTGTTCGGAGGGATCTGATCTGGATAAAGTGCTGCCG 102

427 luArgAlaThrSerValTyrLeuValGlnLysValValProMetLeuPro 443

||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

103 AGAGGGCTACAAAGCGTCTACTTGGTTCAAAAGGTGTCCTCCATCTTCCC 152

444 ArgLeuLeuCysGluGluLeuCysSerLeuAsnProMetThrAspLysLe 460

||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

153 AGGCTGCTGTGTGAGGAGCTGTGCAGCCTCAACCCCATGTCGCAAGCT 202

460 uThrPheSerValIleTrpLysLeuThrProGluGlyLysIleLeuGluG 477

||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

203 GACCTTCTCTGTGATCTGGACACTGACTCCAGAGGGCAAGATCCTGTATG 252

477 luTrpPheGlyArgThrIleIleArgSerCysThrLysLeuSerTyrAsp 493

||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

253 AATGGTTTGGCCGACCATCATCCGCTCCTGCACCAAACTTAGCTACGAG 302

494 HisAlaGlnSerMetIleGluAsnProThrGluLysIleProGluGluG 510

```
|||||
303 CATGCACAGACGATGTTGAAAGCCCAACTGAGAAATCCCTCGAAGA 352
510 uLeuProProHleSerProGluHisSerValGluGluValHisGlnAlav 527
353 GGTGCCCCCATTTCCAGAGCATAGCAGCAGGAGGTACACAGCCG 402
527 aLeuAsnLeuHisSerIleAlaLysGlnLeuArgGlnArgPheVal 543
403 TCTTGAATCTCCACGGAATGCCAAGCAGTTACGCCAGCAGCGTTTGTG 452
544 AspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaPheThrLeuAspHi 560
453 GACGGCGCACTTCTGTTGGATCAGCTTAAAGCTTGTCTTCACTCTGGACCA 502
560 sGluThrGlyLeuProGlnGlyCysHisIleYrGluYrArgAspSera 577
503 CGAGACCGGATGCTCAAGGATGTCATATCTATGAGTACCGCGAGAGCA 552
577 snLysLeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHis 593
553 ACAAGCTCGTGGAGGAGTTCATGCTCTTGCCCAACATGGCAGTGGCCAC 602
594 LysIlePheArgThrPheProGluGlnAlaLeuLeuArgArgHisProPr 610
603 AAGATCCACCGCGCTTCCCGCAGCAGGCCCTGCTGCGCGCACCCCCC 652
610 oProGlnThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetG 627
653 GGCCCAACAGGATGCTCAGTGACCTGTGGAAATCTCGCAGCAGATGG 702
627 lLeuProMetAspValSerSerAlaGlyAla 637
703 GGCTGCGCGTGACTTCAGTCCGCGAGAGCC 734
```

seq\_name: gb\_est2:BF450541

```
seq documentation block:
LOCUS BF450541 681 bp mRNA linear EST 29-DEC-2000
DEFINITION u267f12.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3674159 5'
similar to SW:YR86_CABEL Q09568 HYPOTHETICAL 93.7 KDA PROTEIN
F48E8.6 IN CHROMOSOME III. [2] SW:YR86_CABEL ; mRNA sequence.
ACCESSION BF450541
VERSION BF450541.1 GI:11516710
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 681)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
Clone Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
```

```
MGI:1434927
Seq primer: -40RP from Gibco
High quality sequence stop: 440.
Location/Qualifiers
1. .681
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3674159"
```

FEATURES

source

```
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 160 a 206 c 169 g 144 t 2 others
ORIGIN
alignment_scores:
Quality: 1159,00 Length: 227
Ratio: 5,174 Gaps: 0
Percent Similarity: 98,678 Percent Identity: 98,678
alignment_block:
US-09-647-377-9 x BF450541 ..
Align seg 1/1 to: BF450541 from: 1 to: 681
489 LysLeuSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLys 505
|||||
1 AAACCTGAGCTACGACCATGCCAGCAGCAATGATCGAAAATCCAACTGAGAA 50
505 sileProGluGluGluLeuProProIleSerProGluHisSerValGluG 522
|||||
51 GATCCCTGAGGAAGAGCTTCCGCCAATTTCTCCAGAGCAGCAGCGTCGAGG 100
522 luValHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArg 538
|||||
101 AGGTGCACCGAGCAGTCTCGAACCTGACAGCATGTCGAAAGCAACTCCGC 150
539 ArgGlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAl 555
|||||
151 CGCCAGCGCTTTGTAGATGGCGCACTCCGCTTTAGATCAGCTGAAGCTGC 200
555 aPheThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleYrG 572
|||||
201 TTTTACTCTGGACCATGAGACATGGATTCCTCAAGGATGTCACATCATG 250
572 luTyrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsn 588
|||||
251 AGTACCGAGACGCAACAGCTTGTAGAGGAGTTTCATGCTCTCGGCGCAAC 300
589 MetAlaValAlaHisIlePheArgThrPheProGluGlnAlaLeuLe 605
|||||
301 ATGGCGGTGGCGCCACAGATCTCCGCACCTTCCCTGAGCAGGCGCCTGCT 350
605 uArgArgHisProProGlnThrLysMetLeuSerAspLeuValGluP 622
|||||
351 GGGCGGCATCCCGCCACAGAGATGCTCAGTGACCTGCTGAGGAGGT 400
622 heCysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeu 638
|||||
401 TCTGTGACCAATGGGCTGCCCATGGATGTCAGCTCTGACAGGGGCCCTA 450
639 AsnLysSerLeuThrLysThrPheGlyAspAspLysTyrSerLeuAlaAr 655
|||||
451 AATAAAGCCCTGAGTAAGCATTTGGAGATGACAAAGTACTCTCTGGCCCG 500
655 gLysGluValLeuThrAsnMetTyrSerArgProMetGlnMetAlaLeu 672
|||||
501 GAAGGAGGTGCTCACCACCAATGTTACTCCCGGCCCATGACAGATGCGACTGT 550
672 yPheCysSerGlyMetLeuGlnAspGlnGluInPheArgHisTyrAla 688
|||||
551 ACTTCTGCTCTGGGATGCTCGAGACCAAGACAGTTCGCGCATTAIGCT 600
689 LeuAsnValProLeuTyrThrHisPheThrSerProIleArgArgPheAl 705
|||||
```

601 CTCAAGTGTCCCTTACACACACTTCACCTCTTCCATCCGCGCTTTGC 650  
 705 aaSpValIleValHisArgLeuLeuAlaAla 715  
 651 TGACCTCATAGTGACCGCTTCTGCTGCT 681

seq\_name: gb\_est2:BG864079

seq\_documentation\_block:

LOCUS BG864079 701 bp mRNA linear EST 29-MAY-2001  
 DEFINITION 602797865F1 NIH\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:4918576 5',  
 mRNA sequence.

ACCESSION BG864079  
 VERSION BG864079.1 GI:14214617

KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 701)

REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
 Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10832 row: b column: 17

High quality sequence stop: 701.

FEATURES

Location/Qualifiers

1..701

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:4918576"

/clone\_lib="NIH\_CGAP\_Mam4"

/tissue\_type="tumor, gross tissue"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert 2.5 kb. Library constructed by Life

Technologies, catalog # 12018-016. Investigators providing

Samples: Lothar Hennighausen/Priscilla Furth, NIH

Reference for transgenic model: Li et al., Cell Growth and

Differentiation 7, 3-11 (1996). Note: this is a NCI\_CGAP

Library."

BASE COUNT 171 a 194 c 174 g 162 t

ORIGIN

alignment\_scores:

Quality: 1158.00 Length: 233

Ratio: 5.035 Gaps: 1

Percent Similarity: 98.712 Percent Identity: 96.996

alignment\_block:

US-09-647-377-9 x BG864079 ..

Align seg 1/1 to: BG864079 from: 1 to: 701

328 GluProGluThrGluGlyIleLeuThrGluTyrGlyValAspPheSerAs 344

|||||

2 GAGCCTGAACACAGAGGACTGACAGAAATATGGTGTGACTTCTCTGA 51

|||||

344 pPheSerSerGluValLeuGlySerLeuProGlnSerLeuProTrpThrI 361

|||||

52 TTTCCTCTCAGAGTCTTGAAGTCTCCCTCAAAAGCCTGCCCTGGACAA 101

361 leProProAspGluValGlyLysArgAspLeuArgLysAspCysIle 377  
 |||||  
 102 TCCACACTGATGAGTGGGCAAGAGAACCTTAAGAAAGAGCTGATC 151  
 |||||  
 378 pPheThrIleAspProSerThrAlaArgAspLeuAspAlaLeuAlaC 394  
 |||||  
 152 TTCACCATTTGATCCATCAACTGCTCGGACCTTGATGATGCCCTG 201  
 |||||  
 394 sArgArgLeuThrAspGlyThrPheGluValGlyValHisIleAlaAsp 411  
 |||||  
 202 CAGCGGCTCACTGATGGACCTTTCGAAAGTGGCGTCCACATCGCCG 251  
 |||||  
 411 alSerTyrPheValProGluGlySerSerLeuAspLysValAlaGlu 427  
 |||||  
 252 TGAGTTACTTTGTTCTCTGAGGATCTCTTTGGATAAAGTAGCTG 301  
 |||||  
 428 ArgAlaThrSerValTyrLeuValGlnLysValValProMetLeuPro 444  
 |||||  
 302 AGAGCCACCAAGTGTCTACTTGTCTCAGAAGTGTGTCCTCCCAT 351  
 |||||  
 444 gLeuLeuGlySerGluGluLeuGlySerSerLeuAsnProMetThrAsp 461  
 |||||  
 352 GCTTCTGTGTGAGGAACCTCTCAGCCTCAAGCCATGACTGACAA 401  
 |||||  
 461 hrPheSerValIleThrPysLeuThrProGluGlyLysIleLeuGlu 477  
 |||||  
 402 CCTTCTCTGTGATCTGGAAGCTGACCCCTGAAGGCAAGATCCTT 451  
 |||||  
 478 TrpPheGlyArgThrIleIleArgSerCysThrLysLeuSerTyrAsp 494  
 |||||  
 452 TGGTTTGGCGGCACATATCATCCGTTCTTGCACCAAACTGAGTAG 501  
 |||||  
 494 sAlaGlnSerMetIleGluAsnProThrGluLysIleProGluGlu 511  
 |||||  
 502 TGCCAGACGATGATCGAAATCCACTGAGAGATCCCTGAGGAAAG 551  
 |||||  
 511 euProProIleSerProGluHisSerValGluGluValHisGlnAla 527  
 |||||  
 552 TTCCCCCAATTTCTCCAGAGCACAGCGTCGAGAGGTGCACACGAG 601  
 |||||  
 528 LeuAsnLeuHisSerIleAla..LysGlnLeuArgArgGlnArgPhe 543  
 |||||  
 602 CTGAACCTGCACAGATTGCAAGCCCAACTCCGACGCCAGCGCTGT 651  
 |||||  
 544 AspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaPheThrLeuAsp 559  
 |||||  
 652 AATGGCGCACTCCGTTTATGATCAGTTGAAGCTTGCTTATACTCT 699  
 |||||

seq\_name: gb\_est2:BG746662

seq\_documentation\_block:

LOCUS BG746662 727 bp mRNA linear EST 15-MAY-2001  
 DEFINITION 602703970F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4857353 5',  
 mRNA sequence.

ACCESSION BG746662

VERSION BG746662.1 GI:14057315

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 727)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LLCM1709 row: k column: 18  
High quality sequence stop: 727.

FEATURES

Location/Qualifiers

1. .727  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_15"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
ECORI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 158 a 243 c 194 g 132 t  
ORIGIN

alignment\_scores:  
Quality: 1153.00 Length: 242  
Ratio: 4.970 Gaps: 0  
Percent Similarity: 95.868 Percent Identity: 90.909

alignment\_block:

US-09-647-377-9 x BG746662 ..

Align seg 1/1 to: BG746662 from: 1 to: 727

505 LysIleProGluuGluLeuProIleSerProGluHisSerValGl 521  
||||| :|||  
2 AAAATCCCTGCGAAAGAGCTGCCCCATTTCGCCAGAGCATAGCAGCA 51  
521 uGluValHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuA 538  
52 GGAGGTACACCGCGCTTGAATCTCCACGGAATGCCAAGCATGTAC 101  
538 rgArgGlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeu 554  
||||| :|||  
102 GCCAGCAGGCGCTTTGGACGCGCACTTCGTTGGATCAGCTAAGCTT 151  
555 AlaPheThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTy 571  
152 GCTTTCACTCTGGACCACGAGACCGGATTGCCTCAAGGATGTCATATCTA 201  
571 rGluTyrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaA 588  
202 TGAGTACCGGAGAGCAACAGCTCGTGGAGGAGTTTCATGCTCTTGGCCA 251  
588 snMetAlaValAlaHisLysIlePheArgThrPheProGluGlnAlaLeu 604  
252 ACATGGCAGTGGCCCAACAGATCCACCGCGCTTCCCGCAGCAGGCGCTG 301  
605 LeuArgArgHisProProGlnThrLysMetLeuSerAspLeuValGl 621  
302 CTGGCGCGGACCCCGCCCAAGAGGATCTCAGTGACCTGGTGA 351  
621 uPheCysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaL 638  
352 ATTCGCGACACAGATGGGCTGCCGTTGGACTTCAGCTCCGACGAGGCC 401  
638 euAsnLysSerLeuThrLysThrPheGlyAspAspLysTyrSerLeuAla 654  
402 TCAATAAAGCCCTGACCCCAACATTTGGAGATGACAAGTACTCAGTGGCC 451  
655 ArgLysGluValLeuThrAsnMetTyrSerArgProMetGlnMetAlaLe 671  
452 CGCAAGGAGGTGTCACCAACATGTGCTCCCGGCCCATGCAGATGGCACT 501

671 uTyrPheCysSerGlyMetLeuGlnAspGlnGluInPheArgHisTyrA 688  
||||| :|||  
502 GTACTTCTGTCGGGCTGCTGCAGGACCCAGCGAGTTCGGCAGCTACG 551  
688 laLeuAsnValProLeuTyrThrHisPheThrSerProIleArgArgPhe 704  
552 CGCTCAATGTGCCCTGTACACACACTTCACCTCGCCCATCCGCGCTTT 601  
705 AlaaspValIleValHisArgLeuLeuAlaAlaLeuGlyTyrSerGl 721  
||||| :|||  
602 GCCGACGCTCTGTCGACCGCTCTCTGCTGCCGCTTATAGGGA 651  
721 uGlnProAspValGluProAspThrLeuGlnLysGlnAlaAspHisCysA 738  
652 GCGACTAGACATGCGCGCGCATACCTGCAGAAACAGCGGACCACTGTA 701  
738 snAspArgArgMetAlaSerLysArg 746  
702 ACGACCGCGCATGGCTCCAAAGCG 727

seq\_name: gb\_est2:BM012844

seq\_documentation\_block:

LOCUS BM012844 961 bp mRNA linear EST 30-OCT-2001  
DEFINITION 603637574Fl NIH\_MGC\_47 Homo sapiens cDNA clone IMAGE:5459683 5',  
mRNA sequence.  
ACCESSION BM012844 GI:16527198  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 961)  
TITLE NIH-MGC http://mgs.nci.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1963 row: d column: 20  
High quality sequence stop: 800.

FEATURES

Location/Qualifiers

1. .961  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_47"  
/tissue\_type="neuroblastoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
ECORI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
Note: this is a NIH\_MGC Library."

BASE COUNT 216 a 306 c 255 g 184 t  
ORIGIN

alignment\_scores:

Quality: 1114.50 Length: 300  
Ratio: 4.303 Gaps: 6  
Percent Similarity: 86.333 Percent Identity: 77.667

## alignment\_block:

US-09-647-377-9 x BM012844

Align seg 1/1 to: BM012844 from: 1 to: 961

```
448 GluGluLeuCySerLeuAsnProMetThrAspLysLeuThrPheSerVa 464
|||||
1 GAGGAGCTGTGAGCCTCAACCCCATGTCGACAAAGCTGACCTTCTCTGT 50
464 lletRplysLeuThrProGluGluLysIleLeuGluGluThrPheGlyA 481
|||||
51 GATCTGGACACTGACTCCAGGCGCAAGATCCTTGATGAATGGTTGGCC 100
481 rgThrIleLeuArgSerCysThrLysLeuSerTyrAspHisAlaGlnSer 497
|||||
101 GGACCATCATCGCTCCTCGACCAAACTTAGCTACGAGCATGCACAGC 150
498 MetIleGluAsnProThrGluLysIleProGluGluGluLeuProProII 514
|||||
151 ATGATGTAAGCCCACTGAGAAATCCCTCGAAAGAGCTGCCCCCAT 200
514 eSerProGluHisSerValGluValHisGlnAlaValLeuAsnLeuH 531
|||||
201 TTCCCGCAGCATAGCAGCGAGGAGGTACACAGCGCTCTTGATCTCC 250
531 isSerIleAlaLysGlnLeuArgGlnArgPheValAspGlyAlaLeu 547
|||||
251 AGGGATTTGCCAAGCATTTACGCACGCGCTTTGTGGCGGCGCACTT 300
548 ArgLeuAspGlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLe 564
|||||
301 CGTTTGGATCAGCTAAAGCTTGCTTTCACTCTGGACCACGAGACCGGATT 350
564 uProGlnGlyCysHisIleTyrGlnTyrArgAspSerAsnLysLeuValG 581
|||||
351 GCCTCAAGAGTGTATATCATGTATGATACCGGAGGACACAGCTCGTG 400
581 luGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArg 597
|||||
401 AGGAGTTTCACTGCTTTGGCCAACTGGCAGTGGCCCAAGATCCACCG 450
598 ThrPheProGluGlnAlaLeuLeuArgArgHisProProGlnThrIle 614
|||||
451 GCCTTCCCGGAGCGCCCTGTGCGCGGACACCCCGCCCAACACAG 500
614 sMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeuProMeta 631
|||||
501 GATGCTCAGTGACCTGGTGGAAATTCTGCGACCAAGATGGGCTGCCGTGG 550
631 spValSerSerAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGly 647
|||||
551 ACTTCAGCTCCGAGGAGCCCTCAATAAAGCTGACCAACATTTGGAG 600
648 aspAspLysTyrSerLeuAlaArgLysGluValLeuThrAsnMetTyrS 664
|||||
601 ATGACCAAGTACTACTGGGCGCGGAGGAGGTGCTCACCACATGTGCT 650
664 eArgProMetGlnMetAlaLeu..TyrPheCysSerGlyMetLeuGlnA 680
|||||
651 CCCGCCCATGCAGTGGCACTGTAACTTTCTGCTCGGGCTGCTCGAG 700
680 spGlnGluInPheArgHisTyrAlaLeuAsnValProLeuTyrThrHis 696
|||||
701 ACCAGCGAGTTCGGGCATACCGGCTCAATGTGCCCTGTATTACACAC 750
697 PheThrSerProIleArgArgPheAlaAspValIleValHisArgLeuLe 713
|||||
751 ACTTCCCTCGCCATCGCGCTTTTGGAGGTCCG...GTGCACCGCTCCGTG 797
713 uAlaAlaLeuGlyTyrSerGluInProAspValGluProAspThr. 729
|||||
798 G...TGCCTGTTAGGTATAGGAGCGGACTAAACATGCGCCCG...ATAC 841
```

```
730 LeuGlnLysGlnAlaAspHisCysAsnAspArgArgMetAlaSer 744
|||||
842 CTGCAGAAACAGCGGACCACTGTAAATGACCGCGAGCGTCCAGC 886
```

seq\_name: gb\_est2:BE380437

seq\_documentation\_block:

LOCUS BE380437 725 bp mRNA linear EST 21-JUL-2000  
DEFINITION 601270510F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3599570 5',  
mRNA sequence.

ACCESSION BE380437

VERSION BE380437

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 725)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs@email.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM8782 row: d column: 03

High quality sequence stop: 596.

## FEATURES

Location/Qualifiers

1..725

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:3599570"

/clone\_lib="NCI\_CGAP\_Mam1"

/tissue\_type="tumor, biopsy sample"

/dev\_stage="10 months, virgin"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT 206 a 172 c 179 g 167 t 1 others

ORIGIN

## alignment\_scores:

Quality: 1076.00 Length: 226

Ratio: 4.959 Gaps: 2

Percent Similarity: 96.018 Percent Identity: 93.363

## alignment\_block:

US-09-647-377-9 x BE380437

..

Align seg 1/1 to: BE380437 from: 1 to: 725

146 CysGlyHisHisProLeuGlnGlnSerArgLysGlyTrpSerGlyProAs 162  
|||||

1 TGTGGACATCACCCCTCGACAGTCCCGAAAGGCTGGAGTCTCTGA 50

162 pValIlelleGluAlaGlnPheAspAspSerAspSerGluAspArgHisG 179

|||||

51 TGTCAATTATAGAGCTCAGTTTGTATCAGACGACCTCAGAGATAGACATG 100

179 lyAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerThr 195

|||||

101 GCAACACCAAGTGGCTGGTTGATGATGTTAAGAAATTTGTCATCTCTACT 150

```
196 ProAspGluGlyLysGluAspSerThrProValMetLysAspGluAs 212
|||||
151 CCGACGAGAGGAGAAAGATCTAGTACTCCAGTATGAAGATGAGAA 200
|||||
212 nThrProLeuProGlnAspThrArgGlyLeuSerGluLysSerLeuGlnL 229
|||||
201 CACCCCATACACAGGACACAGAGGCTTATCAGAGAAGTCACTTCAGA 250
|||||
229 ysSerAlaLysValTyrLeuLeuGlyLysHisSerArgAlaAla 245
|||||
251 AATCACAAGGTGTTTACATCTTGAGAGAAAGCAATCTCGACGACCA 300
|||||
246 ThrGlyLeuLeuLysLeuAlaAspLysAsnSerAspLeuPheLysLy 262
|||||
301 ACTGGCATCTCTAAACTCTTGCGCTGATAAGAACAGTGACCTGTTAAGAA 350
|||||
262 sTyrAlaLeuPheSerProSerAspHisArgValProArgIleTyrValP 279
|||||
351 ATACGCCCTGTTTCTCTCTCAGACCAAGGAGTACCTAGAAATTCAGTAC 400
|||||
279 roLeuLysAspCysProGlnAspPheMetThrArgProLysAspPheAla 295
|||||
401 CTCTCAAGGACTGTCGCCAGGACTTCATGACCGACCTAAAGACTTTGCC 450
|||||
296 AsnThrLeuPheLeuCysArgIleLeuAspTyrLysGluAspCysAsnPh 312
|||||
451 ARACGGCTGTTCTCTCGCGCATCATAGACTGGAAGAGGACTGTAAATTT 500
|||||
312 eAlaLeuGlyGlnLeuAlaLysSerLeuGlyGlnAlaGlyGluLleGluP 329
|||||
501 TGCCCTGGGCAACTGGCTAAGAGTCTTGCGCAGGCTGGTGAATTCGAGC 550
|||||
329 roGluThrGluGlyLeuThrGluTyrGlyValAspPheSerAspPhe 345
|||||
551 CTGNAACAGAGGGTACTGACAGATATGGTGGGACTCTCTGATCTC 600
|||||
346 SerSerGluValLeuGluCysLeuProGlnSerLeuProThrThrIlePr 362
|||||
601 TCTTCAGAGCTCTGAATGTC.....TCCTCAGAGCTGCCGGGACAAA 644
|||||
362 opRo...AspGluValGlyLysArgArg 370
|||||
645 TCCACTGATGAGGTGGGACGAGAAAG 671
```

seq\_name: gb\_est2:BG106547

seq\_documentation\_block: 1161 bp mRNA linear EST 30-JAN-2001  
LOCUS BG106547  
DEFINITION 602290354F1 NIH\_MGC\_85 Homo sapiens cDNA clone IMAGE:4385412 5',  
mRNA sequence.

ACCESSION BG106547  
VERSION BG106547.1 GI:12600393  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1161)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue procurement: Louis Staudt, M.D., Ph.D.  
cDNA library preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10066 row: k column: 13  
High quality sequence stop: 639.  
Location/Qualifiers

FEATURES

source 1..1161  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4385412"  
/clone\_lib="NIH\_MGC\_85"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-df primed.  
Average insert size 1.867 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 371 a 291 c 285 g 214 t  
ORIGIN

alignment\_scores:  
Quality: 1072.00 Length: 234  
Ratio: 4.764 Gaps: 1  
Percent Similarity: 96.154 Percent Identity: 86.752  
alignment\_block:  
US-09-647-377-9 x BG106547 ..  
Align seg 1/1 to: BG106547 from: 1 to: 1161  
228 GlnLysSerAlaLysValValTyrIleLeuGluLysLysHisSerArgAl 244  
|||||  
2 CAAAGATCAGCAAGGTGGTTTACATCTTGAGAGAAACATCTCGAGC 51  
|||||  
244 alaThrGlyIleLeuLysLeuAlaAspLysAsnSerAspLeuPheL 261  
|||||  
52 AGCAACCGCGCTTCTCAAACTCTTGCTGATAAGAACAGCAACTGTGTTA 101  
|||||  
261 yslYsTyrAlaLeuPheSerProSerAspHisArgValProArgIleTyr 277  
|||||  
102 GGAATACGCGCTGTTTCTCCTCAGACCCAGCTGCTAGAAATTTAT 151  
|||||  
278 ValProLeuLysAspCysProGlnAspPheMetThrArgProLysAspPh 294  
|||||  
152 GTGCCTCTCAAGGACTGTCCCAGGACTTTGTGGCAGCGCTTAAGATTA 201  
|||||  
294 eAlaAsnThrLeuPheLeuCysArgIleLeuAspTyrLysGluAspCysA 311  
|||||  
202 TGCCAACACACTGTTCATCTGCCCATTTGGGACTGGAAGGAGGACTGCA 251  
|||||  
311 snPheAlaLeuGlyGlnLeuAlaLysSerLeuGlyGlnAlaGlyGluIle 327  
|||||  
252 ATTTTCCCTGGGCGAGCTGGCTAAGAGTCTTGGGCGAGGCTGGTGAAT 301  
|||||  
328 GluProGluThrGluGlyIleLeuThrGluTyrGlyValAspPheSerAs 344  
|||||  
302 GAGCCTGAACAGAGAGGAATACTAACAGAGTATGGCGTGGATTCTCTGA 351  
|||||  
344 pPheSerSerGluValLeuGluCysLeuProGlnSerLeuProThrThrI 361  
|||||  
352 TTTCCTCTCAGAAAGTCTAGAAATGCTTCTCAAGCGCTGCCCATGGACAA 401  
|||||  
361 leProAspGluValGlyLysArgArgAspLeuArgLysAspCysIle 377  
|||||  
402 TTCCACCAGAGGTTTCAGCAAGAGAGGATTTAAGAAAAGACTGTATC 451  
|||||  
378 PheThrIleAspProSerThrAlaArgAspLeuAspAspAlaLeuAlaCy 394  
|||||  
452 TTCACCATTCAGCCCATCAACCGCCGAGACTCGATGATGCGCTCTCCTG 501  
|||||  
394 sArgArgLeuThrAspGlyThrPheGluValGlyValHisIleAlaAspV 411  
|||||  
502 CAGGCCACTCGCTGACGCGCAACTCAAGGTGGGAGTTACATTGCTGAGC 551  
|||||  
411 alSerTyrPheValProGluGlySerSerLeuAspLysValAlaAlaGlu 427  
|||||  
552 TGAGTTTACTTTGTCGGAGGGGATCTGATCTGATCAAGTGGCTGACGAG 601

428 ArgAlaThrSerValTyrLeuValGlnLysValValPrometLeuProAr 444  
 |||||  
 602 AGGCTACAGCGTCTACTTGGTTCACACAGGTGTCCCAATGGCTTCCAG 651  
 |||||  
 444 gLeuLeuCysGluGluLeuCysSerLeuAsnProMetThrAspLysLeu 460  
 |||||  
 652 GCTGCTGTGAAGAACCTGTGCTCTCAACCCAGTTCGAAAGGTG 701  
 |||||

seq\_name: gb\_est2.BG750784

seq\_documentation\_block:

LOCUS BG750784 711 bp mRNA linear EST 15-MAY-2001  
 DEFINITION 602706985F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4843508 5',  
 mRNA sequence.

ACCESSION BG750784  
 VERSION BG750784.1 GI:14061437

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 711)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM1679 row: j column: 21

High quality sequence stop: 702.

FEATURES

source

1..711

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4843508"

/clone\_lib="NIH\_MGC\_43"

/tissue\_type="normal pigmented retinal epithelium"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pOTB7; Site: 1; XhoI; Site 2:

ECORI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC library. |\*

BASE COUNT 168 a 206 c 189 g 148 t

ORIGIN

alignment\_scores:

Quality: 1050.50 Length: 238

Ratio: 4.567 Gaps: 2

Percent Similarity: 96.639 Percent Identity: 90.336

alignment\_block:

US-09-647-377-9 x BG750784 ..

Align seg 1/1 to: BG750784 from: 1 to: 711

406 ValHisLeuAlaAspValSerThrPheValProGluGlnAlaLeu 422

|||||

1 GTTCACATGCTGACGTGAGTACTTTGTCGAGGAGGATCTGATCTGA 50

|||||

422 pLysValAlaAlaGluThrSerValTyrLeuValGlnLysVal 439

|||||

|||||

51 TAAAGTGCTGCCGAGAGGGCTACAAGCGTCTACTTGGTTCAAAAGGTGG 100  
 439 aLPrometLeuProArqLeuLeuCysGluGluLeuCysSerLeuAsnPro 455  
 |||||  
 101 TCCCATGCTTCCAGGCTGCTGTGTGAGGAGCTGTGCAGCCTCAACCC 150  
 |||||  
 456 MetThrAspLysLeuThrPheSerValIleThrLysLeuThrProGlu 472  
 |||||  
 151 ATGTCCGACAAGCTGACCTTCTCTGATCTGTGACACTGACTCCAGAGG 200  
 472 YLysIleLeuGluGluThrPheGlyArgThrIleLeuAspSerCysThrL 489  
 |||||  
 201 CAAGATCCTTGATGATGATGTTTGGCCAGACCATCTCCGCTCTGCACCA 250  
 |||||  
 489 yLeuSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLys 505  
 |||||  
 251 AACTTAGCTACGAGCATGCACAGAGCATGATTGAAGCCCAACTGAGAA 300  
 506 IleProGluGluGluLeuProIleSerProGluHisSerValGluGlu 522  
 |||||  
 301 ATCCCTGCGAAGAGCTGCCGCCCATTTTCCCGAGAGCATAGCAGGAGGA 350  
 |||||  
 522 uValHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgA 539  
 |||||  
 351 GGTACACACAGGCGCTTTGAATCTCCACGGAATTGCCAAGCAGTTACGCC 400  
 539 rGlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAla 555  
 |||||  
 401 AGCAGCGCTTTGGACGGCGCACTTCGTTGGATCAGTAAAGCTTGCT 450  
 |||||  
 556 PheThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrG 572  
 |||||  
 451 TTCACCTCTGGACACGAGACCGGATTCCTCAAGGATGTCATATCTATGA 500  
 572 uTyrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsnM 589  
 |||||  
 501 GTACCGGAGAGCAACAAGCTCGTGGAGGATTCATGCTCTTGCCCAACA 550  
 |||||  
 589 eAlaValAlaHisLysIlePheArgThrPheProGluGlnAlaLeuLeu 605  
 |||||  
 551 TGGCAGTGGC.CACAAGATCCACCGCGCTTCCCGAGCAGGCGCTGCTG 599  
 606 ArgArgHisProPro.ProGlnThrLysMetLeuSerAspLeuValGluP 622  
 |||||  
 600 CGCGCGGCAC...CCCGCGCCCAACAAGATGCTCAGTACCTGGTGGAT 646  
 622 heCysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeu 638  
 |||||  
 647 TCTGCCACAG.ATGGGGCTGCCGCTGGACTTCAGCTCCGCGAGGAGCCTC 695  
 639 AsnLysSerLeu 642  
 |||||  
 696 AATAAAGCGCTG 707

seq\_name: gb\_est2.BF182212

seq\_documentation\_block:

LOCUS BF182212 898 bp mRNA linear EST 31-OCT-2000  
 DEFINITION 601804759F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:4035470 5',  
 mRNA sequence.

ACCESSION BF182212

VERSION BF182212.1 GI:11060354

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 898)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM9309 row: f column: 15  
High quality sequence stop: 656.  
Location/Qualifiers  
1. .898

FEATURES  
source

/organism="Mus musculus"  
/strain="C57/B6"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:4035470"  
/clone\_lib="NCI\_CGAP\_Mam5"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dr.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

BASE COUNT 204 a 232 c 260 g 202 t  
ORIGIN

alignment\_scores:

Quality: 1037.00 Length: 257  
Ratio: 4.339 Gaps: 2  
Percent Similarity: 92.956 Percent Identity: 89.883

alignment\_block:

US-09-647-377-9 x BF182212 ..

Align seg 1/1 to: BF182212 from: 1 to: 898

312 PheAlaLeuGlyGlnLeuAlaLysSerLeuGlyGlnAlaGlyGluIleGl 328  
|||||  
10 TTTGGCTGGGGCACTGGCTAAGAGTCTTGGGCGAGCTGGTGAATCGA 59  
328 uProGluThrGluGlyIleLeuThrGluTyrglyValAspPheSerAsp 345  
60 GCCTGAACAGAGAGGACTGACAGAATATGGTGTGGACTTCTCTGATT 109  
345 heSerSerGluValLeuGluCysLeuProGlnSerLeuProTrpThrIle 361  
|||||  
110 TCCTTCAGAGTTCTTGATGTCTCCCTCAAGCCCTGGCGGCAATC 159  
362 ProProAspGluValGlyLysArgAspLeuArgLysAspCysIlePh 378  
|||||  
160 CCACCTGATGAGTGGGCAAGAGAGACCTTAGGAAAGACTGTATCTT 209  
378 eThrIleAspProSerThrAlaArgAspLeuAspAlaLeuAlaCysA 395  
|||||  
210 CACCATTTGATCCATCACTGCTCGGACCTTGATGATGCCCTGCGTGA 259  
395 r9ArgLeuThrAspGlyThrPheGluValGlyValHisIleAlaAspVal 411  
|||||  
260 GCGGGCTCACTGATGCACCTTCGAAGTGGCGCTCCACATCGCCGATGTG 309  
412 SerTrpPheValProGluGlySerSerLeuAspLysValAlaAlaGluAr 428  
|||||  
310 AGTTACTT.GTTCTTCAGGAGATCCCTC.TTGATATAAGTAGCTCTGAGAG 357  
428 gAlaThrSerValTyrluValGlnLysValValProMetLeuProArgL 445  
|||||  
358 AGCCACAAGTGTCTACTTGGTCCAGAGGTGGTCCCATGCTTCCAGGC 407  
445 euLeuCysGluGluLeuCysSerLeuAsnProMetThrAspLysLeuThr 461  
|||||  
408 TTCGTGTGAGGAACCTCTGCAGCCTCAACCCCATGACTGACAAAGCTGACC 457

462 PheSerValIleTrpLysLeuThrProGluGlyLysIleLeuGluGluTr 478  
|||||  
458 TTTCTCTGTGATCTGGGAAGCTGACCCCTGAAGCGAAGATCTTTGAAGAGTG 507  
478 pPheGlyArgThrIleLeuArgSerCysThrLysLeuSerTyrrAspHisA 495  
|||||  
508 GTTTGGGCGC.ACTATCATCCGTTCTTGACCAAACTGAGCTACGA.CATG 555  
495 laGlnSerMetIleGluAsnProThrGluLysIleProGluGluGluLeu 511  
|||||  
556 CCCAGACATGATCGAAATCAAT.GAGAAGATCCCTGAGGAAGAGCTT 604  
512 ProProIleSerProGluHisSerValGluValHisGlnAlaValLe 528  
|||||  
605 CCCCCAATTTCTCCAGAGCACAGCGTCGAGAGGTGCACCCAGCAGCTCT 654  
528 uAsnLeuHisSerIleAlaLysGlnLeuArgArgGlnArgPheValAspG 545  
|||||  
655 GAACCTGCAGACGATTCGAAAGAAC...TCCGCGCAGAGCGCTTTGATG 701  
545 lyAlaLeuArgLeuAspGlnLeuLysLeuAlaPheThrLeuAspHisGlu 561  
|||||  
702 GCGCACTCGGTTAGATCAGTGAAGCGTTGGGTTTACTCCGCGACCCGG 751  
562 ThrGlyLeuProGlnGlyCys 568  
|||||  
752 GAGGGTTGC...TCAGGGTGT 769

seq\_name: gb\_est2:BI114256

seq\_documentation\_block:

LOCUS BI114256 825 bp mRNA linear EST 26-JUN-2001  
DEFINITION 602862446F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:5021699 5',  
mRNA sequence.

ACCESSION BI114256

VERSION BI114256.1 GI:14565157

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 825)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1CM1840 row: k column: 12

High quality sequence stop: 675.

FEATURES

Location/Qualifiers

1. .825

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="IMAGE:5021699"

/clone\_lib="NIH\_MGC\_17"

/tissue\_type="rhabdomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAGAG(G). Size-selected >500bp

Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies). "

```

BASE COUNT      209 a   235 c   225 g   156 t
ORIGIN
alignment_scores:
  Quality: 1021.00      Length: 275
  Ratio: 4.272          Gaps: 3
  Percent Similarity: 86.909 Percent Identity: 77.818
alignment_block:
US-09-647-377-9 x BI114256 ..
Align seg 1/1 to: BI114256 from: 1 to: 825
394 CysArgArgLeuThrAspGlyThrPheGluValHisIleAlaAs 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3 TGCAGCCACTCGCTGACGGCACTTCAAGTGGAGTGCATGCTGA 52
410 pValSerTyPheValProGluGlySerSerLeuAspLysValAlaAG 427
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
53 CGTGAGTTACTTTGTTCCGGAGGATCTGATCTGGATAAAGTGCTGCCG 102
427 luArgAlaThrSerValTyLeuValGlnLysValValProMetLeuPro 443
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 AGAGGCTACAAAGCTCTACTTGGTTCAAAAGGTGGTCCCATGCTTCCC 152
444 ArgLeuLeuCysGluGluLeuLeuCysSerLeuAsnProMetThrAspLysLe 460
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 AGGCTGCTGTGTGAGGAGCTGTGCAGCCTCAACCCCATGTCGACAAGCT 202
460 uThrPheSerValIleTyPlysLeuThrProGluGlyLysIleLeuGluG 477
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 GACCTTCTCTGTGATCTGGACACTGACTCCAGAGGGCAAGATCCTTGATG 252
477 luThrPheGlyArgThrIleIleArgSerCysThrLysLeuSerTyAsp 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
253 AATGGTTTGGCGGACCATCATCGCTCCGACCAAACTTAGTACGAG 302
494 HisAlaGlnSerMetIleGluAsnProThrGluLysIleProGluGluG 510
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303 CATGCACAGAGCATGATTGAAAGCCCAACTGAGAAAATCCCTGCGAAAGA 352
510 uLeuProTyrLeuSerProGluHisSerValGluGluValHisGlnAlav 527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
353 GCTGCCCCCATTTCCCCAGAGCATAGCAGGAGGAGGTACCGGCCG 401
527 alLeuAsnLeuHisSerIleAlaLysGlnLeuArgArgGlnArgPheVal 543
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
402 TCTTGAATCTCCACGGAATTGCCAAGCAGTTACGCCAGCAGCGCTT.GTG 450
544 AspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaPheThrLeuAspHi 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
451 GACGGCGCACTTCGTCTGGATCAGCTAAAGCTTGCTTCACTCTGGACCA 500
560 sGluThrGlyLeuProGlnGlyCysHisIleTyArgAspSerA 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 CGAGACCGGATTCGCTCAAGGATGTCATATCTAGTACCGCGAGAGCA 550
577 snLysLeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHis 593
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
551 ACAAGCTCTGGAGGAGTTTCACTCTTGCCCAACATGGCAGTGGCCAC 600
594 LysIlePheArgThrPheProGluGlnAlaLeuLeuArgArgHisProPr 610
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 AAGATCCACCGCCCTTCCCGAGCAGGCGCTGCATGCGCGGACACCCC 650
610 oProGlnThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetG 627
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
651 GGCCCCAACAGAGATGCTCAGTACCTAGGGGAATCTGCGCACCAGATG 700
627 lyLeuProMetAspValSerSerAlaGlyAlaLeuAsnLysSerLeu... 642
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 GG.....GCTGGCGGTGGAAACAACAGCATACGG 729

```

```

643 .....ThrLysThrPheGlyAsp.Asplyst 651
|||||:|||||:|||||:|||||:|||||:|||||:|||||
730 AGGAACACTACATAATAAAGCCAGGACACAAACATTAGGAGATTGACAAAG 779

```

```

651 yrSerLeuAlaArgLysGluVal 658
|||||:|||||:|||||:|||||:|||||:|||||:|||||
780 ACTCACTAGGCGCAGAGAGGTG 802

```

seq\_name: gb\_est2:BI988256

```

seq_documentation_block:
LOCUS      BI988256                600 bp      mRNA      linear      EST 20-DEC-2001
DEFINITION 3223-16 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,

```

ACCESSION BI988256

VERSION BI988256.1 GI:17959235

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 600)

AUTHORS Mu X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S. W.,

White, R. A., Beremand, P. D., Thomas, T. L., Gan, L., and Klein, W. H.,

gene expression in the developing mouse retina by EST sequencing

and microarray analysis

Nucleic Acids Res. 29 (24), 4983-4993 (2001)

JOURNAL Contact: Klein WH

COMMENT Department of Biochemistry and Molecular Biology

University of Texas M.D. Anderson Cancer Center

Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA

Tel: 713 792 3646

Fax: 713 790 0329.

FEATURES Location/Qualifiers

source

1..600

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

/tissue\_type="neural retina"

/dev\_stage="embryonic day 14.5 post-fertilization"

BASE COUNT 143 a 171 c 149 g 137 t

ORIGIN

alignment\_scores:

Quality: 1019.00

Ratio: 5.121

Percent Similarity: 99.500

Percent Identity: 99.500

alignment\_block:

US-09-647-377-9 x BI988256 ..

Align seg 1/1 to: BI988256 from: 1 to: 500

401 ThrPheGluValGlyValHisIleAlaAspValSerTyPheValProGl 417

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

2 ACCTTCAAGAGTGGCGTCCACATCGCGATGTGAGTTACTTTGTTCTGTA 51

417 uGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyL 434

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

52 GGATCTCTTTGGATAAAGTAGCTGCTGAGAGAGCCACAAAGTGTCTACT 101

434 euValGlnLysValValProMetLeuProArgLeuLeuCysGluGluLeu 450

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

102 TGGTCCAGAAAGGTGGTCCCATGCTTCCAGGCTTCTGTGTGAGGAACCT 151

451 CysSerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLy 467

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

152 TGCAGCTCAACCCATGACTGACACAGCTGACCTTCTCTGTGATCTGAA 201

467 LeuThrProGluGlyLysIleLeuGluThrPheGlyArgThrIleI 484

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

202 GCTGACCCTGAAGCAAGATCTTGAAGAGTGGTTTGGCGCAGCTATCA 251  
 484 leArgSerCysThrLysLeuSerTyAspHisAlaGlnSerMetIleGlu 500  
 |||||||  
 252 TCCGTTCTTGCACCAAACTGAGCTAGCAGATGCCACGAGCATGATCGA 301  
 |||||||  
 501 AsnProThrGluLysIleProGluGluGluLeuProProIleSerProGl 517  
 |||||||  
 302 AATCCAATGAGAAGATCCCTGAGGAAGAGCTTCCCCCAATTCTCCAGA 351  
 |||||||  
 517 uHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleA 534  
 |||||||  
 352 GCACAGCGCTGAGGAGGTGCACGAGCAGTCTGAACCTGCACAGCATTG 401  
 |||||||  
 534 laLysGlnLeuArgGlnArgPheValAspGlyAlaLeuArgLeuAsp 550  
 |||||||  
 402 CAAGCAACTCCGCGCCAGCGCTTGTAGATGGCGCACTCCGTTAGAT 451  
 |||||||  
 551 GlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnGl 567  
 |||||||  
 452 CAGCTGAAGCTTGCTTTTACTCTGGACCATGAGACTGGATTGCCCTCAAG 501  
 |||||||  
 567 yCysHisIleTyArgAspSerAsnLysLeuValGluGluPhe 583  
 |||||||  
 502 ATGTACATCTATGATACCGGAGACAGCAACAAGCTTGTAGAGAGATT 551  
 |||||||  
 584 MetLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhe 599  
 |||||||  
 552 ATGCTCTGGCCACATGGCGGTGGCCCAAGATCTTCCGCACTTC 599

seq\_name: gb\_est2:BG804017

seq\_documentation\_block:  
 LOCUS BG804017 600 bp mRNA linear EST 20-DEC-2001  
 DEFINITION 0244-55 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,

ACCESSION BG804017  
 VERSION BG804017  
 KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 600)  
 AUTHORS Mu.X., Zhao.S., Pershad.R., Hsieh.T.-F., Scarpa.A., Wang.S.W.,  
 White.R.A., Beremand.P.D., Thomas.T.L., Gan.L. and Klein.W.H.

TITLE Gene expression in the developing mouse retina by EST sequencing  
 and microarray analysis

JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
 COMMENT Contact: Klein WH

Department of Biochemistry and Molecular Biology  
 University of Texas M.D. Anderson Cancer Center  
 Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
 Tel: 713 792 3646  
 Fax: 713 790 0329.

FEATURES Location/Qualifiers

source  
 1..600  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Mouse E14.5 retina lambda ZAP II Library"  
 /tissue\_type="neural retina"  
 /dev\_stage="embryonic day 14.5 post-fertilization"  
 BASE COUNT 147 a 173 c 151 g 126 t 3 others  
 ORIGIN

alignment\_scores:  
 Quality: 1015.00 Length: 200  
 Ratio: 5.152 Gaps: 1  
 Percent Similarity: 98.500 Percent Identity: 98.500

alignment\_block:  
 US-09-647-377-9 x BG804017

Align seg 1/1 to: BG804017 from: 1 to: 600

448 GluGluLeuCysSerLeuAsnProMetThrAspLysLeuThrPheSerVa 464  
 |||||||  
 2 GAGGAATCTGCAAGCTNNAACCCATGACTGACAAAGCTGACCTTCTCTGT 51  
 |||||||  
 464 lileTrpLysLeuThrProGluGlyLysIleLeuGluGluTrpPheGlyA 481  
 |||||||  
 52 GATCTGGAAGCTGACCCCTGAAGGCAAGATCCTTGAAGAGTGGTTTGCC 101  
 |||||||  
 481 rgThrIlelleArgSerCysThr.LysLeuSerTyAspHisAlaGlnSe 497  
 |||||||  
 102 GCATATCATCCGTTCTTGCACCAANTCTTGAGCTAGCAGCATGCCACGAG 151  
 |||||||  
 497 rMetIleGluAsnProThrGluLysIleProGluGluGluLeuProProI 514  
 |||||||  
 152 CATGATCGAAAAATCCAACCTGAGAAGATCCCTGAGGAAGAGCTTCCCCAA 201  
 |||||||  
 514 leSerProGluHisSerValGluGluValHisGlnAlaValLeuAsnLeu 530  
 |||||||  
 202 TTTCTCCAGAGCACAGCGTCGAGGAGGTGCACCGAGCAGTCTGAAACCTG 251  
 |||||||  
 531 HisSerIleAlaLysGlnLeuArgGlnArgPheValAspGlyAlaLe 547  
 |||||||  
 252 CACAGCATTCGAAAGCAACTCCGCCCGCCAGCGCTTTGTAGATGGCGCACT 301  
 |||||||  
 547 uArgLeuAspGlnLeuLysIleAlaPheThrLeuAspHisGluThrGlyL 564  
 |||||||  
 302 CCGTTAGATCAGCTGAAGCTTGTCTTTACTCTGGACCATGAGACTGGAT 351  
 |||||||  
 564 euProGlnGlyCysHisIleTyArgGlnArgAspSerAsnLysLeuVal 580  
 |||||||  
 352 TGCCTCAAGGATGTCACATCTATGATACCGAGACAGCAACAAGCTTGT 401  
 |||||||  
 581 GluGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysIlePheAr 597  
 |||||||  
 402 GAGGAGTTTCATGCTCTGGCCCAACATGGCGGTGGCCCAAGATCTCCG 451  
 |||||||  
 597 gThrPheProGluGlnAlaLeuLeuArgHisProProGlnThrL 614  
 |||||||  
 452 CACCTTNCCTGAGCAGCGCTCTGTCGCCGCGCATCCGCCACACAGACGA 501  
 |||||||  
 614 ysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeuPromet 630  
 |||||||  
 502 AGATCTCAGTCACTGCTGGTGGAGTTCTGTGACCAAGTGGGCTGCCCATG 551  
 |||||||  
 631 AspValSerAlaGlyAlaLeuAsnLysSerLeuThrLysThrPhe 646  
 |||||||  
 552 GATGTACGCTCTGCAGGGGCCCTAAATAAAGCCTGACTAAGACATTT 599

seq\_name: gb\_est2:BE284271

seq\_documentation\_block:

LOCUS BE284271 622 bp mRNA linear EST 26-OCT-2000  
 DEFINITION 601087339F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3482313 5',  
 mRNA sequence.

ACCESSION BE284271

VERSION BE284271.1 GI:9161093

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 622)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cyagbs-r@mail.nih.gov](mailto:cyagbs-r@mail.nih.gov)

Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM8511 row: n column: 10  
 High quality sequence stop: 569.

Location/Qualifiers  
 1. .622

#### FEATURES

source  
 /organism="Mus musculus"  
 /strain="C57/B6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3482313"  
 /clone\_lib="NCI\_CGAP\_Mam5"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;  
 Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH"

BASE COUNT 152 a 172 c 155 g 143 t

#### ORIGIN

#### alignment\_scores:

Quality: 1015.00 Length: 206  
 Ratio: 5.025 Gaps: 2  
 Percent Similarity: 98.058 Percent Identity: 96.117

#### alignment\_block:

US-09-647-377-9 x BE284271 ..

Align seg 1/1 to: BE284271 from: 1 to: 622

333 GlyIleLeuThrGluTyrGlyValAspPheSerAspPheSerGluVa 349

|||||

3 GGGATACGACGAAATATGTGGGACTTCTGATTCTTCAGAAAT 52

|||||

349 lleuGluCysLeuProGlnSerLeuProTrpThrIleProAspGluV 366

|||||

53 TCTTGAATGTCTCCCTCAAGGCTGCCTGGACAATCCACCGTGATGAGG 102

|||||

366 aGlyLysArgArgAspLeuArgLysAspCysIlePheThrIleAspPro 382

|||||

103 TGGCGACGAGACGAGACCTAAGAAAGACTGTATCTTCACCATTTGATCA 152

|||||

383 SerThrIaArgAspLeuAspAlaLeuAlaCysArgArgLeuThrAs 399

|||||

153 TCAACTGCTCGGACCTTGATGATGCCCTGCCTGCAGCGGCTCACTCA 202

|||||

399 pGlyThrPheGluValGlyValHisIleAlaAspValSerTyrPheValP 416

|||||

203 TGGCACCTTCGAAGTGGGCTCCACATCGCGATGTGAGTTACTTTGTTTC 252

|||||

416 roGluGlySerSerLeuAspLys.ValAlaAlaGluArgAlaThrSerVa 432

|||||

253 CTGAGGGATCCTCTTTGGATAACAGTAGCTGCTGAGAGAGCCCAAGTGT 302

|||||

432 lTyrlleuValGlnLysValProMetLeuProArgLeuLeuCysGluG 449

|||||

303 CTACTTGGTCAGAGTGGTCCCATGCTTCCAGGCTTCTGTGTGAGG 352

|||||

449 luLeuCysSerLeuAsnProMetThrAspLysLeuThrPheSerValIle 465

|||||

353 AACTCTGACGCTCAACCCCATGACTGACAAGCTGACCTTCTCTGTGATC 402

|||||

466 TrpLysLeuThrProGluClyLysIleLeuCluIleTrpPheClyArgTh 482

|||||

403 TGGAGCTGACCCCTGAAGGCAAGATCCTTGAAGAGTGGTTTGGCGCAC 452

|||||

482 rIleIleArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetI 499

|||||

453 TATCATCCGTTCTTGTGCACAACTAGCTAGCACCAGCATGCCAGACATGA 502

|||||

499 leGluAsnProThrGluLysIlePleProGluGluGluLeuProProIleSer 515

|||||

503 TCGAAATCCAACTGAGAAGATCCCTGTGAGGAGAGCTTCCCCCAATTCT 552

|||||

516 ProGluHisSerValGluClu.ValHisGlnAlaValLeuAsnLeuHis 532

|||||

553 CCAGACACAGCTCGAAGAGGTGCACAGGCGAGTCTCTGAACCTGCACAG 602

|||||

532 erlLeAlaLysGln 536

|||||

603 CATTGCCAAGGCAA 616

|||||

seq\_name: gb\_est2:BI650394

seq\_documentation\_block:

BI650394 1158 bp mRNA linear EST 12-SEP-2001

DEFINITION 603296125F1 NIH\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5336719 5',

mRNA sequence.

ACCESSION BI650394

VERSION BI650394.1 GI:15564630

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 1158)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgaps-femail.nih.gov](mailto:cgaps-femail.nih.gov)

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM1855 row: a column: 08

High quality sequence stop: 762.

Location/Qualifiers

1. .1158

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:5336719"

/clone\_lib="NIH\_CGAP\_Mam3"

/tissue\_type="tumor, gross tissue"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: NotI;

Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert 2 kb. Library constructed by Life

Technologies, catalog #12017-018. Investigators providing

samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference

for transgenic model: Xu et al., Nature Genetics 22, 37-43

(1999). Note: this is a NCI\_CGAP Library."

BASE COUNT 351 a 259 c 326 g 221 t 1 others

ORIGIN

#### alignment\_scores:

Quality: 1002.50 Length: 259  
 Ratio: 4.321 Gaps: 9  
 Percent Similarity: 89.575 Percent Identity: 83.784

#### alignment\_block:

US-09-647-377-9 x BI650394 ..

Align seg 1/1 to: BI650394 from: 1 to: 1158

1 MetAsnHisProAspTyrLysLeuAsnLeuArgSerProGlyThrProAr 17

|||||

```
131 ATGAACCATCTGCTACTACAGCTGACCTTCGGTCTCGGGACCCCCAG 180
17 gGlyValSerSerValValGlyProSerAlaValGlyAlaSerProGlyA 34
|||||
181 AGGTGTGTCCTCTGTGTTGGCCGAGTCTCTTGGTCTTCGCCAGGTG 230
34 sPLysLysSerLysAsnLysSerMetArgGlyLysLysLysSerIlePhe 50
|||||
231 ACARAAAGTCAAGAACAAGTCCATCGGAGGAGGAGAAAGAGCATATTT 280
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67
|||||
281 GAAACCTACATGTCCAAGGAGGATGTTTCAGAAAGCTTGAAGAGAGAAC 330
67 rLeuIleGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84
|||||
331 ACTTATCCAGGCTGATTGAGANTCAACCCAAAGAAAGTTTCATGAAGCCT 380
84 heIleProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100
|||||
381 TCATTCTCTTCCTCCGGATGGTGATCGGGACATTTTATTGATGGAGTTGTT 430
101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuPr 117
|||||
431 GCTCGTANTAGAGCCTTAATGGGACCTTGTGGTTGTAAACTGCTTCC 480
117 oGluaspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIleG 134
|||||
481 TGAGGATCAGTGGGAGGACGTTAAACACAGAGCAATGACAAAGAAATAG 530
134 luAlaThrTyrGluAlaAspIleProGluGluGlyCysGlyHisHisPro 150
|||||
531 AGCTACTATTGAGCTGACATCCCTGAAGAGGGCTGTGGACATCACCC 580
151 LeuGlnGlnSerArgLysGlyTrpSer..GlyProAspValIleIleGlu 166
|||||
581 CTGCACAGTCCCGGAAAGGCTGGAGTTGGTCTCTGATGTCATTATAG 630
167 AlaGlnPheAspSerAspSerGluAspArgHisGlyAsnThrSerGl 183
|||||
631 GCTTCAGTTGATGACAGGACTCAGAAGATAGACATGGCAACACCATGG 680
183 yLeuValAspGlyValLysLysLeuSerIleSer...ThrProAspArg 199
|||||
681 CTGTGTGATGGTGTTCGACACACTTGTCAATCTCTAACTCCTCTGACAG 730
199 lLysGlu..AspSerSerThr.ProValMetLys.AsuGluAsnThrPr 214
|||||
731 GCARAGACAGATTCTAGTACTCCAGTTATGACAGATGAGACCCACC 780
214 o...IleProGlnAspThrArgGlyLeuSerGluLys.....S 226
|||||
781 CCATATACAAAGGCCACAGAGCTTATCCAGAGGAGTCCCTCTCTG 830
226 erLeuGlnLysSerAlaLysValValTyrIleLeuGluLysLysHis 242
:::|||||
831 AACTTCAGCAA.....CGGTGGGTACACATCTTGAGAGAGAAAGAGTT 874
243 ArgAla...AlaThrGlyIle 248
:::|||||
875 CTCGACAGGCAACTGGCATC 895
```



OM of: US-09-647-377-9 to: GenEmbl.\* out\_format : pfs

Date: Jul 8, 2002 2:11 PM

About: Results were produced by the GenCore software, version 4.5,

Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=cgn2\_1/USPTO\_spool/US09647377/runat\_08072002\_080021\_23069/app\_query.fasta\_1.940  
-DB=GenEmbl -OFT=Fastap -SUFFIX=olip2n.rge -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000 -GAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=60.000 -GAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR\_SCORE=quality  
-THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09647377 @CGN1\_112153 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-647-377-9

Query length: 870

Database: GenEmbl.\*

Database sequences: 1797656

Database length: 1873333701

Search time (sec): 2783.320000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
Sequence					
gb_pat:AX015905	+	780.00	16983.40	3100	! AX015905 Sequence 8 from Pat
gb_pat:AX015899	+	202.00	3914.26	1550	! AX015899 Sequence 1 from Pat
gb_ro:AF312994	+	83.00	1551.62	4.6e-78	228283 ! AF312994 Mus musculus chrom
gb_pat:AX015902	+	78.00	1463.99	3.5e-73	49999 ! AX015902 Sequence 5 from Pat
gb_pat:AX015908	+	78.00	1463.99	3.5e-73	49999 ! AX015908 Sequence 11 from Pat
gb_pr:AB066536	+	63.00	1191.49	5.3e-58	2189 ! AB066536 Macaca fascicularis
gb_pat:AX015903	+	61.00	1133.32	9.2e-55	36901 ! AX015903 Sequence 6 from Pat
gb_pat:AX015909	+	61.00	1132.97	9.6e-55	38886 ! AX015909 Sequence 12 from Pat
gb_pat:AC098189	-	61.00	1123.71	3.1e-54	153701 ! AC098189 Rattus norvegicus
gb_pat:AC024062	+	40.00	712.32	2.6e-31	162842 ! AC024062 Homo sapiens chrom
gb_pat:AF312225	+	40.00	707.26	4.9e-31	345102 ! AF312225 Homo sapiens chrom
gb_pat:AX015907	+	39.00	700.70	1.1e-30	49999 ! AX015907 Sequence 10 from Pat
gb_pat:AX015910	+	36.00	664.44	1.2e-28	1784 ! AX015910 Sequence 13 from Pat
gb_pat:AC098189	+	36.00	634.42	5.6e-27	153701 ! AC098189 Rattus norvegicus
gb_pr:AC019130	+	30.00	517.67	1.8e-20	139049 ! AC019130 Homo sapiens BAC cl
gb_pat:AC105461	-	30.00	514.58	2.7e-20	219775 ! AC105461 Homo sapiens chrom
gb_pat:AC024062	+	26.00	438.32	4.7e-16	162842 ! AC024062 Homo sapiens chrom
gb_pat:AX015901	+	25.00	452.17	8.0e-17	1140 ! AX015901 Sequence 3 from Pat
gb_pat:AX015904	+	25.00	435.60	6.7e-16	13330 ! AX015904 Sequence 7 from Pat
gb_pat:AX015913	+	25.00	426.70	2.1e-15	49999 ! AX015913 Sequence 16 from Pat
gb_pat:AC093374	+	25.00	418.24	6.2e-15	175391 ! AC093374 Homo sapiens chrom
gb_pat:AC068134	+	25.00	418.03	6.4e-15	181068 ! AC068134 Homo sapiens BAC cl
gb_pat:AC068134	+	19.00	309.27	7.3e-09	49999 ! AX015912 Sequence 15 from Pat
gb_pat:AC105461	+	18.00	288.12	1.1e-07	63173 ! AC105461 Mus musculus clone R
gb_pat:AC105461	+	18.00	279.72	3.2e-07	219775 ! AC105461 Homo sapiens chrom
gb_pat:AX015914	+	16.00	250.55	1.4e-05	49999 ! AX015914 Sequence 17 from Pat
gb_pat:AC092106_4	+	15.00	239.19	5.8e-05	14789 ! Continuation (5 of 5) of AC09
gb_in:u23514	+	15.00	233.32	0.0001	35325 ! u23514 Caenorhabditis elegans
gb_pat:AX015916	+	15.00	230.98	0.0002	49999 ! AX015916 Sequence 19 from Pat
gb_pat:AC012033	+	15.00	223.22	0.0005	158276 ! AC012033 Homo sapiens clone
gb_pat:AC074188	+	15.00	222.36	0.0005	179816 ! AC074188 Homo sapiens chrom
gb_pat:AF307337	+	15.00	221.31	0.0006	210301 ! AF307337 Homo sapiens ALPP,
gb_in:AY057479	+	12.00	191.24	0.0274	3992 ! AY057479 Arabidopsis thaliana
gb_in:CEC04G2	+	12.00	174.90	0.2224	23830 ! 270718 Caenorhabditis elegans
gb_pl:AC007584	-	12.00	168.92	0.4789	82189 ! AC007584 Arabidopsis thaliana
gb_pl:AC012193	-	12.00	168.90	0.4802	82454 ! AC012193 Arabidopsis thaliana
gb_pat:AC098739	+	12.00	168.52	0.5040	87197 ! AC098739 Oryza sativa chromos
gb_pl:AP000615	-	12.00	164.68	0.8244	154128 ! AP000615 Oryza sativa genom

gb\_ba:AF269873 + 11.00 170.43 0.3943 3592 ! AF269873 Staphylococcus epi  
gb\_pat:AX145191 - 11.00 170.43 0.3943 3592 ! AX145191 Sequence 3913 from  
gb\_htg:PFMAL13P2\_2 - 11.00 147.38 7.58 110000 ! Continuation (3 of 4) of  
em\_htg:inv:AC057824 + 10.00 160.63 1.39 842 ! AC057824 Giardia intestinali  
gb\_pr:AK001346 + 10.00 157.07 2.19 1429 ! AK001346 Homo sapiens cDNA  
gb\_pr:HS0800675 + 10.00 154.08 3.21 2228 ! AL080158 Homo sapiens mRNA;  
gb\_pr:AK025634 + 10.00 153.80 3.33 2324 ! AK025634 Homo sapiens cDNA;

seq\_name: gb\_pat:AX015905

seq\_documentation\_block:

LOCUS AX015905 3100 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 8 from Patent WO9950284.  
ACCESSION AX015905  
VERSION AX015905.1 GI:10041648

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3100)

Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.

Nucleic acid molecules which code proteins influencing bone

development

Patent: WO 9950284-A 8 OCT-1999;

aigner THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP

ANDREAS (DE); WIRTH THOMAS (DE)

Location/Qualifiers

1..3100

/organism="Mus musculus"

/db\_xref="taxon:10090"

125..2737

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC07620.1"

/db\_xref="GI:10041649

/translation="MNHPTKLNLSPTGPGVSSVVGPSAVGASPGDKSKNKSMDRG

KKSIFFETYSKEDVSEGLKRTGLQVLRINPKKHEAFIPSPGDRDIFIDGVVAR

NKALGDLVVKLLPEDQWKAVKPESNDKEATEYDIEPEGCGHHPQQRKGWSG

PDIIVTAQFDSDSDEHNTSGLDVGKLSISTDPRGKEDSTPMKDNTPIDQD

TRGLSEKSLQKSAKVVIILEKHSRAATGILKLLADKNSDLFKKALFSPSDHVPRI

VYPLKDPQDMTPRDKFANTLCRIIDWKEDCNFALGQKSLGQAGEISPETEGI

LTEYGVDFSPFSEVLECLPSQLPTIPDDEVGRRLRDKCIFTDSTARDLDAAL

ACRLTDTGTEVGVHDAVSYVPEGSSLDKVAERATSVLVKVPMLPRLCEEL

CSLNPMDTKLFFSVYINKLIPESGKILEEFWFGTITRSCTKLSYDHAQSMLENTEKPE

ELPPIPSHSEVVEEHOVLNLSAKORQRFVQALRRHPPQTKMLSDLVLEPCD

GCHYIEYRDSNKLVEFMFLANNAVAHIFRFEPEQALLRRHPPQTKMLSDLVLEPCD

QMLPMDVSSAGALNLSLTQDDKQISLARKEVITNNYSRPMALYFCSGMLQDQE

OPRHVALNPPLYTHFTSPTRRFADVIHRLAAALGYSEQDVEPDTLQKQADHCNDR

RMASKRVOELISGLFFAVLVKESGLESEAMVMGVLMQAFDLVLRFGVKRYICNAL

ALRSYFQKVGKKPELTLYWEPDLEETQQTITFSLDVVLQAEATALKYSAILK

RPLKESASDEEPED"

BASE COUNT 783 a 795 c 812 g 710 t

ORIGIN

alignment\_scores:

Quality: 870.00 Length: 870

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x AX015905

Align seg 1/1 to: AX015905 from: 1 to: 3100

1 MetAsnHisProAspTyrLysLeuAsnLeuArgSerProGlyThrProAr 17

|||||

125 ATGACCACTCTCTGACTACAGCTGACCTTCGGTCCGGGAGCCCCAG 174

|||||

17 gGlyValSerValGlyProSerAlaValGlyAlaSerProGlyA 34

|||||

175 AGGTGTGCTCTGCTGTGGTGGCCGAGTGTCTGTGTGCTTCAGGTG 224

|||||

34 spLysLysSerLysAsnLysSerMetArgGlyLysLysLysSerLysPhe 50  
 |||||  
 225 ACAAAAGTCAAGAGCAAGTCCATCGAGGGAGCAAGAAAGAGCATATTT 274  
 |||||  
 51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67  
 |||||  
 275 GAAACCTACATGTCCAGAGGAGATGTTTCAGAAAGCTTGAAGAGAGAAC 324  
 |||||  
 67 rLeuLeuGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84  
 |||||  
 325 ACTTATCAGGGGTGTATTGAGATCAACCAAGAAAGTTTCATCAAGSCT 374  
 |||||  
 84 heileProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100  
 |||||  
 375 TCATTCCCTTCCCGATGGGTATCGGGACATTTTATTGATGGAGTTGTT 424  
 |||||  
 101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuPr 117  
 |||||  
 425 GCTCTAATAGAGCCTTAATGGGACCTTGTGGTTGTAAACTGCTTCC 474  
 |||||  
 117 oGluAspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIleG 134  
 |||||  
 475 TGAGGATCAGTGGAGGCGATTAAACGAGAGCAATGACAAAGAAATAG 524  
 |||||  
 134 luAlaThrTyrGluAlaAspIleProGluGluGlyCysGlyHisHisPro 150  
 |||||  
 525 AAGCTTACTTATGAGCTGACATCCCTGAAGAGGGCTGTGGACATCACCC 574  
 |||||  
 151 LeuGlnGlnSerArgLysGlyTrpSerGlyProAspValIleIleGluAl 167  
 |||||  
 575 CTGACGAGCTCCCGAANGCTGGAGTGGTCTGATGTCAITATAGAGGC 624  
 |||||  
 167 acLlnPheAspSerAspSerGluAspArgHisGlyAsnThrSerGlyL 184  
 |||||  
 625 TCAGTTTGTATGAGCGGACTCAGAAGATAGACATGGCAACACCATGGCC 674  
 |||||  
 184 euValAspGlyValLysLysLeuSerIleSerThrProAspArgGlyLys 200  
 |||||  
 675 TGGTTGATGGTGTAAAGAAATTGTCAATCTCTACTCTGACAGAGGAAA 724  
 |||||  
 201 GluAspSerThrProValMetLysAspGluAsnThrProIleProGlu 217  
 |||||  
 725 GAAGATTCTAGTACTCCAGTTATGAAGATGAGAACAACCCCATACACA 774  
 |||||  
 217 nAspThrArgGlyLeuSerGluLysSerLeuGlnLysSerAlaLysValV 234  
 |||||  
 775 GGACACAAGAGGCTTATCAGAGAAGTCACTTCAGAAATCAGCAAGGTGG 824  
 |||||  
 234 alTyrIleLeuGluLysLysHisSerArgAlaAlaIleThrGlyLleLeuLys 250  
 |||||  
 825 TTTACATCTTGGAGAAAAGCAATCTCGAGCAGCAACTGGCATTCCTGAAA 874  
 |||||  
 251 LeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaLeuPheSe 267  
 |||||  
 875 CTCTTGGCTGATAAAGACAGTGAACCTGTTTAAGAAATACGCCCTGTTTC 924  
 |||||  
 267 rProSerAspHisArgValProArgIleTyrValProLeuLysAspCysP 284  
 |||||  
 925 TCCTTCAGACCACCGAGTACCTAGAAATTACGTACCTCTCAAGGACTGTC 974  
 |||||  
 284 roGlnAspPheMetThrArgProLysAspPheAlaAsnThrLeuPheIle 300  
 |||||  
 975 CCCAGGACTTCATGACCGACCTAAAGACTTTGCCAACACGCTGTTCATC 1024  
 |||||  
 301 CysArgIleIleAsnTrpLysGluAspCysAsnPheAlaLeuGlyGlnLe 317  
 |||||  
 1025 TGCCTCATATAGACTGGAGAGGAGCTGTAATTTTGCCTTGGGCAACT 1074  
 |||||  
 317 uAlaLysSerLeuGlyGlnAlaGlyGluIleGluProGluThrGluGlyI 334  
 |||||  
 1075 GGCTAAGAGTCTTGGCAGGCTGTGAAATCGAGGCTGAAACAGAGAGGGA 1124  
 |||||

334 leLeuThrGluTyrGlyValAspPheSerAspPheSerSerGluValLeu 350  
 |||||  
 1125 TACTGACAGAAATATGGTGGACTTCTCTGATTTCTCTTCAGAAAGTTCTT 1174  
 |||||  
 351 GluCysLeuProGlnSerLeuProTrpThrIleProProAspGluValG1 367  
 |||||  
 1175 GAATGTCTCCTCAAGGCTGCCCTGGACAATCCCACCTGATGAGGTGGG 1224  
 |||||  
 367 yLysArgArgAspLeuArgLysAspCysIlePheThrIleAspProSerT 384  
 |||||  
 1225 CAGAGAAAGAGACCTAAGAAAGACTGTATCTTCACCATTTGATCCATCAA 1274  
 |||||  
 384 hrAlaArgAspLeuAspAspAlaLeuAlaCysArgArgLeuThrAspGly 400  
 |||||  
 1275 CTGCTCGCAGCTTGATGATGCCCTGCCCTGCAGGCGCTCACTGATGGC 1324  
 |||||  
 401 ThrPheGluValGlyValHisIleAlaAspValSerTyrPheValProG1 417  
 |||||  
 1325 ACCTTCGAAGTGGGCTCCACATCGCGATGTGAGTTACTTTGTTCCCTGA 1374  
 |||||  
 417 uGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrL 434  
 |||||  
 1375 GGGATCCTCTTTGGATAAAGTAGCTGCTGAGAGAGCCACAAGTGTCTACT 1424  
 |||||  
 434 euValGlnLysValValProMetLeuProArgLeuLeuCysGluGluLeu 450  
 |||||  
 1425 TGGTCCAGAGGTGTCCCATGCTTCCCAGGCTCTGTGTGAGGAACCTC 1474  
 |||||  
 451 CysSerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLy 467  
 |||||  
 1475 TGCAGGCTCAACCCATGACTGACAAAGCTGACCTTCTGTGATCTGGAA 1524  
 |||||  
 467 sLeuThrProGluGlyLysIleLeuGluGluTrpPheGlyArgThrIleI 484  
 |||||  
 1525 GCTGACCCCTGAAGCGCAAGATCCTTGAAGAGTGGTTGGCCGCACTATCA 1574  
 |||||  
 484 leArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGlu 500  
 |||||  
 1575 TCGTTTCTTGCACCAAACTGAGCTACGACATGCCAGCATGCCAGCATGATCGAA 1624  
 |||||  
 501 AsnProThrGluLysIleProGluGluGluLeuProProIleSerProG1 517  
 |||||  
 1625 AATCCAACCTGAGAAGATCCCTGAGGAAGAGCTTCCCCCAATTTCCAGA 1674  
 |||||  
 517 uHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleA 534  
 |||||  
 1675 GCACAGCGTCCGAGGAGGTGCACAGCGAGTCTGSAACCTGCACAGCATTG 1724  
 |||||  
 534 laLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAsp 550  
 |||||  
 1725 CAAAGCAACTCCGCCGCCAGCGCTTGTAGATGGCGCACTCCGTTTAGAT 1774  
 |||||  
 551 GlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnG1 567  
 |||||  
 1775 CAGCTGAAGCTTGCTTTTACTCTGGACCATGAGATGGACTGGCTTCAAGG 1824  
 |||||  
 567 yCysHisIleTyrGluTyrArgArgSerAsnLysLeuValGluGluPheM 584  
 |||||  
 1825 ATGTCACATCTATGAGTACCGAGACAGCAACAAGCTTGTAGAGGAGTTCA 1874  
 |||||  
 584 etLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhePro 600  
 |||||  
 1875 TGCTCCTGGCCACATGGCGGTGGGCCCAAGATCTTCCGCACTTCCCT 1924  
 |||||  
 601 GluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSe 617  
 |||||  
 1925 GAGCAGGCCCTTGCTGGCGCGCATCCCCACACAGAGGAAGATGCTCAG 1974  
 |||||  
 617 rAspLeuValGluPheCysAspGlnMetClyLeuProMetAspValSerS 634  
 |||||  
 1975 TGACTGTGTGAGTTCTGTGACCCAGATGGGCTGCCATGGATGTGAGCT 2024  
 |||||  
 634 erAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspLys 650  
 |||||

```
|||||
2025 CTCGAGGGCCCTAAATAAAGAGCTGACTAAGACATTTGGAGATGACAAG 2074

651 TyrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMe 667
|||||
2075 TACTCTCTGGCCCGAAGAGGTGCTACCAACATGTAATCTCCGGCCCAT 2124

667 tGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluNP 684
|||||
2125 GCAGATGGCACTGTACTTCTGCTCTGGGATGCTGCAGACCAAGAGCAGT 2174

684 heArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerPro 700
|||||
2175 TCCGGCATATATGCTCTCAAGTTCCCTCTACACACATTTACACTCTCC 2224

701 IleArgGArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLe 717
|||||
2225 ATCCGCGCTTGTGCTGACGTCAATGTCACCGCTCTGCTGCTCTCT 2274

717 uGlyTyrSerGluGlnProAspValGluProAspThrLeuGlnLysGlnA 734
|||||
2275 GGGCTACAGTGAACAGCCAGATGTGGAGCTGTATACCTACAGAGCAAG 2324

734 laAspHisCysAsnAspArgMetAlaSerLysArgValGlnGluLeu 750
|||||
2325 CTGACCACTGCAATGACCGTGCATGCTTCCAAACGTTGCAGAGGCTC 2374

751 SerIleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuG 767
|||||
2375 AGCATTGGGCTCTTCTCGCAGTTCTAGTAAGAGGAGTGGCCCCCTGGA 2424

767 uSerGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuV 784
|||||
2425 GTCCGAAGCCATGGTGATGGGTGCTCTGAACCAAGCTTTCGACGTGCTGG 2474

784 alLeuArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeu 800
|||||
2475 TGCTGGCGCTTTGGGTGCAGAAGCGCATCTACTGCAATGCACCTGAG 2524

801 ArgSerTyrSerPheGlnLysValGlyLysLysProGluLeuThrLeuVa 817
|||||
2525 CGATCTACAGTTCCGAAGGTGGGAAGAGGAGCAACACAGAGGTCAATCA 2574

817 lTrpGluProAspAspLeuGluGluProThrGlnGlnValIleThrI 834
|||||
2575 TTGGAGCCTGATGATACCTTTGAAGAGGAGCAACACAGAGGTCAATCA 2624

834 lePheSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLys 850
|||||
2625 TCTTCAGCTGGTGGATGTTGCTTGCAGGAGAGGCCACAGCCCTCAAG 2674

851 TyrSerAlaIleLeuLysArgProGlyLeuGluLysAlaSerAspGluG 867
|||||
2675 TACAGTGCTATCTGAAGCGACCGAGGCTGGAGAAGCGCTCTGATGAGGA 2724

867 uProGluAsp 870
|||||
2725 GCCTGAGGAC 2734

seq_name: gb_pat:AX015899
seq_documentation_block:
LOCUS AX015899 1550 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from Patent WO9950284.
ACCESSION AX015899
VERSION AX015899.1 GI:10041641
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1550)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
```

```
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 1 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Location/Qualifiers
1. 1550
/organism="Mus musculus"
/db_xref="taxon:10090"
<2..1183
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAC07618.1"
/db_xref="GI:10041642"
/translation="LGRSKVAARATISVILVQKVPMLRPLRLLCEELCSLNPWTDKLTFF
SVIWLTPGKILLEWFGRTIIRSCITKLSYDHAQSMTEPTKIPPEELPPTISPEHSV
FEVHQAIVLNLSIAKOLRRQFVDGALRDOEFMLLANMAVAHKIFRTFPQALLRRH
PPQTKMLSLDVEFCQDGLPMDYSAGALNMLYPCSGMLQDQOFQFRHYALNPVLYT
HFTSPIRRFADIVYHRLAALGALGSEQDPDPTLOKQADHCNDRMASHKSVQELSIG
LFFAVLVKESGPLESEAMVGVNQAFLVLRFGVOKRIYCNALALASYSFQKVGKK
PELTLVWEPDDLEETQOVIITIFSLVDVLQAEATALKYSAILKRPGLERASDEEPE
D"
BASE COUNT 352 a 433 c 412 g 353 t
ORIGIN
```

```
alignment_scores:
Quality: 202.00 Length: 202
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-647-377-9 x AX015899
Align seg 1/1 to: AX015899 from: 1 to: 1550
669 MetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluGlnPheAr 685
575 ATGGCACTGTACTTCTGCTGGGATGCTGAGGACCAAGAGGAGGAGTCCG 624
685 gHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIlea 702
625 GCATTATGCTCTCAACGTTCCCTCTACACACACTTCACTCTCCCATCC 674
702 rgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGly 718
675 GCCGCTTTGCTGAGTCATAGTCACCGCTCTGCTGCTGCTGCTGCTGGGC 724
719 TyrSerGluGlnProAspValGluProAspThrLeuGlnLysGlnAlaAs 735
725 TACAGTGAACAGCCAGATGTGGAGCCTGATACCTACAGAGCAAGCTGA 774
735 pHIScysAsnAspArgMetAlaSerLysArgValGlnGlnLeuSerI 752
775 CCACCTGCAATGACCGTCGATGCTTCCAAAGCTGTGCAGGAGGAGCTCAGCA 824
752 leGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuGluSer 768
825 TCGGCTCTTCTTCGAGTTCTAGTAAGAGAGAGTGGCCCTCGAGTCC 874
769 GluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuVal 785
875 GAAGCCATGGTATGGGTGCTGCTGAACCAAGCTTTCGAGCTGCTGGTGT 924
785 uArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArg 802
925 GCGCTTTGGGTGCAGAAGCGCATCTACTGCAATGCACTGCGCCCTCGCAT 974
802 erTyrSerPheGlnLysValGlyLysProGluLeuThrLeuValTrp 818
975 CCTACAGCTTCCAGAAGTGGGGAAGCAAGCAGAGCTCACTCTGTGTGG 1024
819 GluProAspAspLeuGluGluProThrGlnGlnValIleThrIlePh 835
```

```

|||||
1025 GAGCCTGATGACCTTGAGAGGAGCCACACAGCAGGTCATCACCATCTT 1074
|||||
835 eSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLysTyrS 852
|||||
1075 GAGCCTGGTGATGTGGTCTGAGGAGGCCACAGCCCTCAAGTACA 1124
|||||
852 eRalAlaLeuLysArgProGluLeuGluLysAlaSerAspGluGluPro 868
|||||
1125 GTGCTATCTCTGAAGCAGCAGGCGCTGGAGAAGCGCTCTGATGAGGACCT 1174
|||||
869 GluAsp 870
|||||
1175 GAGGAC 1180

seq_name: gb_ro:AF312994

seq_documentation_block:
LOCUS AF312994 228283 bp DNA linear ROD 17-JAN-2002
DEFINITION Mus musculus chromosome 1 clone MML, complete sequence.
ACCESSION AF312994
VERSION AF312994.1 GI:18182288
KEYWORDS FIG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Rump,A., Hess,J., Galsocz,P., Wirth,T. and Rosenthal,A.
TITLE Genomic sequence from mouse chr. 1
JOURNAL Unpublished
REFERENCE
AUTHORS Rump,A., Hess,J., Galsocz,P., Wirth,T. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
Location/Qualifiers
source 1..228283
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="MML"

BASE COUNT 63688 a 49622 c 49154 g 65819 t
ORIGIN

alignment_scores:
Quality: 83.00 Length: 83
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AF312994 ..
Align seg 1/1 to: AF312994 from: 1 to: 228283

233 ValValTyrIleLeuGluLysLysHisSerArgAlaAlaThrClyIleLeu 249
|||||
58690 GTGGTTTACATCTGTGGAGAAAGCATTTCTGAGCAGCAACTGGCATCCT 58739
|||||
249 uLysLeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaLeuP 266
|||||
58740 GAAACTCTGGCTGATAGAAGACAGTGACCTGTTTAAAGAAATACGCCCTGT 58789
|||||
266 heSerProSerAspHisArgValProArgIleTyrValProLeuLysAsp 282
|||||
58790 TTTCCTCTCAGACACCGAGTACTAGAATTTACGTACTCTCTCAAGGAC 58839
|||||
283 CysProGlnAspPheMetThrArgProLysAspPheAlaAsnThrLeuP 299
|||||
58840 TGTCGCCAGGACTTCATGACCGGACCTTAAGACTTTGCCACACACGCTGTT 58889
|||||
299 eIleCysArgIleLeuAspTrpLysGluAspCysAsnPheAlaLeuGly 315
|||||

```

```

|||||
58980 CATCTGCCGATCATAGACTGGAAGAGGAGCTGTAAATTTTGCCCTGGGG 58938
|||||
seq_name: gb_pat:AX015902

seq_documentation_block:
LOCUS AX015902 49999 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 5 from Patent WO950284.
ACCESSION AX015902
VERSION AX015902.1 GI:10041645
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 5 07-OCT-1999;
ALGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Location/Qualifiers
source 1..49999
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 13210 a 11814 c 10825 g 14150 t
ORIGIN

alignment_scores:
Quality: 78.00 Length: 78
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AX015902 ..
Align seg 1/1 to: AX015902 from: 1 to: 49999

474 IleLeuGluGluTrrPheGlyArgThrIleIleArgSerCysThrLysLe 490
|||||
8520 ATCTTTGAAGAGTGGTTGGCCGACATCATCTCCGTTCTTGACCAACT 8569
|||||
490 uSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysIleP 507
|||||
8570 GAGCTACCAACCATGCCAGAGCATGATCGAAAATCCAACTGAGAAGATCC 8619
|||||
507 roGluGluGluLeuProProlSerProGluHisSerValGluGluVal 523
|||||
8620 CTGAGGAAGAGCTTCCCCCAATTTCTCCAGAGCACAGCGTCGAGAGGTG 8669
|||||
524 HisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArgL 540
|||||
8670 CACCAAGGAGTCTGAAACCTGCACAGCATTCGAAAGCAACTCCGCCGCCA 8719
|||||
540 nArgPheValAspGlyAlaLeuArgLeuAspGln 551
|||||
8720 GCGCTTTGTAGATGCGCACTCCGCTTTAGATCAG 8753
|||||

seq_name: gb_pat:AX015908

seq_documentation_block:
LOCUS AX015908 49999 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 11 from Patent WO950284.
ACCESSION AX015908
VERSION AX015908.1 GI:10041651
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 49999)

```

AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.  
 TITLE Nucleic acid molecules which code proteins influencing bone development  
 JOURNAL PATENT: WO 9950284-A 11 07-OCT-1999;  
 ATGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP ANDREAS (DE); WIRTH THOMAS (DE)  
 FEATURES Location/Qualifiers  
 source  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 BASE COUNT 13135 a 11787 c 10868 g 14209 t  
 ORIGIN

alignment\_scores:  
 Quality: 78.00 Length: 78  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-647-377-9 x AX015908 ..  
 Align seg 1/1 to: AX015908 from: 1 to: 49999

474 lleLeuGluGluTrpPheGlyArgThrIleIleArgSerCysThrLysLe 490  
 |||||  
 10505 ATCTTGAAGAGTGGTTTGGCGGCACATCATCCGTTCTTGACCAAACT 10554  
 |||||  
 490 userTyAspHisAlaGlnSerMetIleGluAsnProThrGluLysIleP 507  
 |||||  
 10555 GACTACAGACCATGCCAGACATGATCGAATCACTGAGAAGATCC 10604  
 |||||  
 507 roGluGluLeuProProlleSerProGluHisSerValGluGluVal 523  
 |||||  
 10605 CTGAGGAAGAGTTCCTCCCAATTTCTCCAGAGCAGCGTCGAGGAGTG 10654  
 |||||  
 524 HisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArgL 540  
 |||||  
 10655 CACCAGGCAGTCCTCAACCTGCACAGCATTTGCAAGCAACTCCGCGCCA 10704  
 |||||  
 540 nArgPheValAspGlyAlaLeuArgLeuAspGln 551  
 |||||  
 10705 GCGCTTGTAGATGGCGCACTCCGTTTAGATCAG 10738  
 |||||

seq\_name: gb\_pr:AB066536

seq\_documentation\_block:  
 LOCUS AB066536 2189 bp mRNA linear PRI 26-JUL-2001  
 DEFINITION Macaca fascicularis brain cDNA clone:QmoA-12105, full insert sequence.  
 ACCESSION AB066536  
 VERSION AB066536.1 GI:15021873  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Macaca fascicularis adult male medulla oblongata cDNA to mRNA, clone.lib:macaque brain cDNA library QmoA clone:QmoA-12105.  
 ORGANISM Macaca fascicularis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopithecinae; Macaca.  
 1 (bases 1 to 2189)  
 Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.  
 Isolation of full-length cDNA clones from macaque brain cDNA libraries  
 REFERENCE 1 (bases 1 to 2189)  
 AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JUL-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
 (E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)

seq\_name: gb\_pat:AX015903

seq\_documentation\_block:  
 LOCUS AX015903 36901 bp DNA linear PAT 07-SEP-2000  
 DEFINITION Sequence 6 from Patent WO9950284.  
 ACCESSION AX015903  
 VERSION AX015903.1 GI:10041646  
 KEYWORDS house mouse.

COMMENT Lab host: TOP10  
 Vector: pME18S-FL3 (ACC.No. AB009864)  
 R. Site1: DraIII (CACCATGTCG)  
 R. Site2: DraIII (CACCATGTCG)  
 Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGCGCTTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection PCR product was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing ( 5' end primer [CTTCTGCTCTAAAGCTGG]; 3' end primer [CGACCTGCAGCTCGAGCACA] ).

FEATURES  
 Location/Qualifiers  
 1..2189  
 /organism="Macaca fascicularis"  
 /db\_xref="taxon:9541"  
 /clone="QmoA-12105"  
 /sex="male"  
 /tissue\_type="medulla oblongata"  
 /clone\_lib="macaque brain cDNA library QmoA"  
 /dev\_stage="adult"  
 151..900  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="BAB62212.1"  
 /db\_xref="GI:15021874"  
 /translation="MSHPDYRMNLRPLGTGRGVSTVAGPHGIGASPGDKKSKNKSTRGKKKSIFETVMFKEDVSEGLKRGTLQIGVLRINPKRPFHAFIPSPDGRDRIFDGVVARNRALNGLVVVKLLPEQOMKVKVPENKDETEAAAYESDIPEELCGRHLPOQLSKSYNDSPDVTIEAOFDSGSDSEHGHTONVLDVVKKLSVCVSEKRGREHGDAPVTKDETTCTISQDTRALPEKSLQRSKAVIAVRFSSHVQMALI"  
 BASE COUNT 603 a 419 c 522 g 645 t  
 ORIGIN

alignment\_scores:  
 Quality: 63.00 Length: 63  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-647-377-9 x AB066536 ..  
 Align seg 1/1 to: AB066536 from: 1 to: 2189

56 LysGluAspValSerGluGlyLeuLysArgGlyThrLeuIleGlnGlyVa 72  
 |||||  
 316 AAGGAGGATGTTTCGGAAGCGCTTGAAGAGAGGAGCACTCATCCAGGTGT 365  
 |||||  
 72 lleuArgIleAsnProLysLysPheHisGluAlaPheIleProserProA 89  
 |||||  
 366 ATTGAATTAATCAAGAAGTTTCATGAAGCCTTCATCTTCCCGG 415  
 |||||  
 89 spGlyAspArgAspIlePheIleAspGlyValValAlaArgAsnArgAla 105  
 |||||  
 416 ATGTTGATCGAGACATTTTATTGATGGGTGTTGCTCGTAATAGACC 465  
 |||||  
 106 LeuAsnGlyAspLeuValValLysLeuLeuProGlu 118  
 |||||  
 466 TTAATGGGGATCTGTTGGTGGTGAAGAACTACTTCCCGAG 504  
 |||||

```

ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          1 (bases 1 to 36901)
Nucleic acid molecules which code proteins influencing bone
development
JOURNAL        Patent: WO 9950284-A 6 07-OCT-1999;
AUTHOR        AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES       Location/Qualifiers
source         1..36901
               /organism="Mus musculus"
               /db_xref="taxon:10090"
BASE COUNT    9597 a 9409 c 8973 g 8922 t
ORIGIN
alignment_scores:
  Quality:    61.00      Length:    61
  Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-647-377-9 x AX015903 ..
Align seg 1/1 to: AX015903 from: 1 to: 36901
579  LeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysIle 595
12478 CTTGTAGAGAGTTCATGCTCTGCGCCACATGCGGTGGCCCAAGAT 12527
595  ePheArgThrPheProGluGlnAlaLeuLeuArgHisProProProG 612
12528 CTTCCGCACCTTCCTGCTGAGCAGGCCCTGCTGCGCGGCATCCCCACCAC 12577
612  InThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeu 628
12578 AGACGAGAGTCTCAGTGCCTGGGAGTCTGTGTGACGAGTGGGGCTG 12627
629  ProMetAspValSerSerAlaGlyAlaLeuAsn 639
12628 CCCATGGATGTCAGCTCTGCAGGGGCCCTAAAT 12660
seq_name: gb_pat:AX015909
seq_documentation_block:
LOCUS      AX015909
DEFINITION Rattus norvegicus clone WO9950284. linear PAT 07-SEP-2000
ACCESSION AX015909
VERSION    AX015909.1 GI:10041652
KEYWORDS   house mouse.
SOURCE     Mus musculus.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 38886)
Nucleic acid molecules which code proteins influencing bone
development
JOURNAL    Patent: WO 9950284-A 12 07-OCT-1999;
AUTHOR     AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES   Location/Qualifiers
source     1..38886
           /organism="Mus musculus"
           /db_xref="taxon:10090"
BASE COUNT 10087 a 9920 c 9356 g 9523 t
ORIGIN
alignment_scores:
  Quality:    61.00      Length:    61

```

```

Ratio: 1.000      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-647-377-9 x AX015909 ..
Align seg 1/1 to: AX015909 from: 1 to: 38886
579  LeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysIle 595
14462 CTTGTAGAGAGTTCATGCTCTGCGCCACATGCGGTGGCCCAAGAT 14511
595  ePheArgThrPheProGluGlnAlaLeuLeuArgHisProProProG 612
14512 CTTCCGCACCTTCCTGCTGAGCAGGCCCTGCTGCGCGGCATCCCCACCAC 14561
612  InThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeu 628
14562 AGACGAGAGTCTCAGTGCCTGGGAGTCTGTGTGACGAGTGGGGCTG 14611
629  ProMetAspValSerSerAlaGlyAlaLeuAsn 639
14612 CCCATGGATGTCAGCTCTGCAGGGGCCCTAAAT 14644
seq_name: gb_htg:AC098189
seq_documentation_block:
LOCUS      AC098189
DEFINITION Rattus norvegicus clone HT230-145D14. *** SEQUENCING IN PROGRESS
ACCESSION AC098189
VERSION    AC098189.3 GI:17973688
KEYWORDS   HTG; HTGS. PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
           1 (bases 1 to 153701)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burch,P., Burkett,C., Burrell,K.D., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,B., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Ieal,B., Lewis,D.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,M., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

```



\* 140943 142718: contig of 1776 bp in length  
 \* 142719 142818: gap of unknown length  
 \* 142819 143980: contig of 1162 bp in length  
 \* 143981 144080: gap of unknown length  
 \* 144081 145323: contig of 1243 bp in length  
 \* 145324 145423: gap of unknown length  
 \* 145424 146688: contig of 1265 bp in length  
 \* 146689 146788: gap of unknown length

alignment\_scores:  
 Quality: 61.00 Length: 61  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-647-377-9 x AC098189/rev ..

Align seg 1/1 to reverse of: AC098189 from: 1 to: 153701

579 LeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysI 595

77874 CTGTGAGGAGGTTCATGCTCTGGCCACATGGCAGTGGCCACAGAT 77825

595 ePheArgThrPheProGluGlnAlaLeuLeuArgHisProProG 612

77824 CTTCGGACCTTCCTGAGCAGGCCCTGCTGCGCGGCATCCCCACCG 77775

612 InThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeu 628

77774 AGACGAGATGCTCAGTGACCTGGTGAGTCTGTGACCATGATGGGCGTG 77725

629 ProMetAspValSerSerAlaGlyAlaLeuAsn 639

77724 CCCATGGATGTCAGCTCTGCAGGGGCCCTAAAT 77692

seq\_name: gb\_htg:AC024062

seq\_documentation\_block:

LOCUS AC024062 162842 bp DNA linear HTG 10-MAR-2001  
 DEFINITION Homo sapiens chromosome 2 clone RP11-95H14, WORKING DRAFT SEQUENCE,  
 7 unordered pieces.

ACCESSION AC024062

VERSION AC024062.27 GI:13270576

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Abola,A.P., Bruno,D., Conn,L., Della Rosa,M., Faulkner,D.,  
 Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,  
 Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,  
 Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,  
 Yu,S. and Davis,R.W.

JOURNAL

AUTHORS

Unpublished  
 2 (bases 1 to 162842)  
 Bruno,D., Conn,L., Della Rosa,M., Faulkner,D., Federspiel,N.,  
 Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,  
 Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,  
 Yu,S. and Davis,R.W.

TITLE

JOURNAL

Submitted (22-FEB-2000) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,

USA

On Mar 10, 2001 this sequence version replaced gi:13236644.

----- Genome Center

Center: Stanford DNA Sequencing and Technology Development

Center

Center code: SDSTDC

Web site: <http://sequence-www.stanford.edu/group/human/>

Contact: [hum.info@sequence.stanford.edu](mailto:hum.info@sequence.stanford.edu)

----- Project information

Center project name: 842

Center clone name: RP11-95H14

----- Summary Statistics

Sequencing Vector: M13mp18; X02513; 99% of reads

Chemistry: Dye-terminator; plasmid; plasmid\_accession; 1% of reads

Chemistry: Dye-terminator; 2% of reads

Chemistry: Dye-terminator Big Dye; 97% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 159366 bases at least Q40

Consensus quality: 159909 bases at least Q30

Consensus quality: 160434 bases at least Q20

Insert size: 166936; agarose-fp

Insert size: 162242; sum-of-contigs

Quality coverage: 14.6x in Q20 bases; agarose-fp

Quality coverage: 15.0x in Q20 bases; sum-of-contigs.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 7 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1779: contig of 1779 bp in length

\* 1780 1879: gap of unknown length

\* 1880 11374: contig of 9495 bp in length

\* 11375 11474: gap of unknown length

\* 11475 38208: contig of 26734 bp in length

\* 38209 38308: gap of unknown length

\* 38309 65551: contig of 27243 bp in length

\* 65552 65651: gap of unknown length

\* 65652 93881: contig of 28230 bp in length

\* 93882 127838: contig of 33857 bp in length

\* 127839 127938: gap of unknown length

\* 127939 162842: contig of 34904 bp in length.

## FEATURES

Location/Qualifiers

1..162842

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/clone\_lib="RPC1 human BAC library 11"

1..1779

/note="assembly\_name:Contig59"

1880..11374

/note="assembly\_name:Contig61"

11475..38208

/note="assembly\_name:Contig62"

38309..65551

/note="assembly\_name:Contig63"

clone\_end:SF6"

65652..93881

/note="assembly\_name:Contig64"

93982..127838

/note="assembly\_name:Contig65"

clone\_end:T7"

127939..162842

/note="assembly\_name:Contig66"

BASE COUNT 45634 a 33939 c 33845 g 48820 t 604 others

ORIGIN

## alignment\_scores:

Quality: 40.00 Length: 40

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-647-377-9 x AC024062 ..

Align seg 1/1 to: AC024062 from: 1 to: 162842

316 GlnLeuAlaLysSerLeuGlyGlnAlaGlyGlnLeuGluProGluThrG1 332

```

|||||
79514 CAGCTGGCTAAGAGCTTGGCGCAGGCTGGTGAATTGAGCCTGAAACAGA 79563
|||||
332 uGlytLeuThrGluTyrGlyValAspPheSerAspPheSerGluV 349
|||||
79564 AGGATACTACAGAGTATGGCTGGATTCTCTGATTCTCTTCACAGAG 79613
|||||
349 alLeuGluCysLeuProGln 355
|||||
79614 TTCTAGAATGCTCTCTCA 79633
|||||

seq_name: gb_pr:AF312225
seq_documentation_block:
LOCUS AF312225 345102 bp DNA linear PRI 17-JAN-2002
DEFINITION Homo sapiens chromosome 2 map 2q37.1, complete sequence.
ACCESSION AF312225
VERSION AF312225.1 GI:18182287
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 345102)
AUTHORS Rump,A., Hess,J., Wen,G., Wirth,T. and Rosenthal,A.
TITLE Human genomic sequence from 2q37.1
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 345102)
AUTHORS Rump,A., Hess,J., Wen,G., Menzel,U., Wirth,T. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
Location/Qualifiers
1..345102
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37.1"

BASE COUNT 92011 a 70334 c 74209 g 108499 t 49 others
ORIGIN

alignment_scores:
Quality: 40.00 Length: 40
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AF312225 ..
Align seg 1/1 to: AF312225 from: 1 to: 345102
316 GlnLeuAlaLysSerLeuGlyGlnAlaGlyGluLeuGluProGluThrG 332
178531 CAGCTGGCTAAGAGCTTGGCGCAGGCTGGTGAATTGAGCCTGAAACAGA 178580
332 uGlytLeuThrGluTyrGlyValAspPheSerAspPheSerGluV 349
178581 AGGAATACTAACAGAGTATGGCGTGATTCTCTGATTCTCTTCACAGAG 178630
349 alLeuGluCysLeuProGln 355
178631 TTCTAGAATGCTCTCTCA 178650
|||||

seq_name: gb_pat:AX015907
seq_documentation_block:
LOCUS AX015907 49999 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 10 from Patent WO9950284.
ACCESSION AX015907
VERSION AX015907.1 GI:10041650
KEYWORDS house mouse.
SOURCE

```

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 49999)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 10 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
source
Location/Qualifiers
1..49999
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 14550 a 10214 c 10719 g 14516 t
ORIGIN

alignment_scores:
Quality: 39.00 Length: 39
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AX015907 ..
Align seg 1/1 to: AX015907 from: 1 to: 49999
400 GlyThrPheGluValGlyValHisIleAlaAspValSerTyrPheValPr 416
30110 GGCACCTTCGAAGTGGCGTCACATCGCGATGTGAGTTACTTTGTGCC 30159
416 cGluGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValT 433
30160 TCAGGGATCTCTTTGGATTAAGTAGCTGCGAGAGCCACAGGTCT 30209
433 yrLeuValGlnLysVal 438
30210 ACTTGGTCCAGAGGTA 30226
|||||

seq_name: gb_pat:AX015910
seq_documentation_block:
LOCUS AX015910 1784 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 13 from Patent WO9950284.
ACCESSION AX015910
VERSION AX015910.1 GI:10041653
KEYWORDS human.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1784)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 13 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
source
Location/Qualifiers
1..1784
/organism="Homo sapiens"
/db_xref="taxon:9606"
<3..1454
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC07621.1"
/db_xref="GI:10041654"
/translaton="NFKVGVHIAVDSYFVPEGSDDLKVAARATSYLVQKVVPMPLP
R
TEKEELSLNPMSDKLTFSVIWTLPTEGKILDEWFGRTIIRSCTKLSYSHAQSMIESP
LTKIPAKELPPISEHSSEVHOAVNLNLGTAQQRQRFVDGLAKDQIKLAFTLDH
ETGLPQGCCHIYEYRESKNLVEEFMLANMAVAHKIHRAPTEQALLRRHPPQTRMLSD
LVEFCDQMGLPVDFSSAGALNLSLTQTTFGDDKYSLARKEVLTNCSRPMQALYFCSG

```

```
LLQDPAQFRHYALNPVLYTHFTSPIRRFADVLVHLAAALGYRERLDMAPDTLQKQA
DHCNDRMRASKRVOELSTSLFFAVLVKESGPLESEAWMGILIKQADVLVLRGVQKR
IYCNALALRSHHFQVKGKPELTLVNEPDMRQEPAAQOVITIFSLVENVLQAEALTK
YSAILKRPCTGHLGPKPEEESDGPEDSSTS*
BASE COUNT      385 a   337 c   304 g   358 t
ORIGIN
```

```
alignment_scores:
  Quality:      36.00      Length:      36
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment\_block:

US-09-647-377-9 x AX015910 ..

Align seg 1/1 to: AX015910 from: 1 to: 1784

421 LeuAspLysValAlaLaGluAArgAlaThrSerValTyrLeuValGlnLys 437

63 CTGGATAAGTGGCTCCGAGAGGCTACAGCGTCTACTTGGTTCARAA 112

437 sValValProMetIeuProArgIeuLeuCysGluGluLeuCysSerLeuA 454

113 GGTGTCCTCCCATGCTTCCACAGGCTGCTGTGTGAGGAGCTGTGACGCTCA 162

454 snPromet 456

163 ACCCCATG 170

seq\_name: gb\_htg\_AC098189

seq\_documentation\_block:

LOCUS AC098189 153701 bp DNA linear HTG 20-DEC-2001  
DEFINITION Rattus norvegicus clone CH230-145D14, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 61 unordered pieces.

ACCESSION AC098189.3 GI:17973688

VERSION AC098189.3

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 153701)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,

Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,D., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,B., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,B., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,

Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,

Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokan,I., Rolfe,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,  
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczyk,K., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 153701)

Worley,K.C.

Direct Submission

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17064556.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GIHF

Center clone name: CH230-145D14

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to  
findPhrapList

Consensus quality: 123669 bases at least Q40

Consensus quality: 129396 bases at least Q30

Consensus quality: 135066 bases at least Q20

Estimated insert size: 124661; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 61 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 8487: contig of 8487 bp in length

8487: gap of unknown length

8587: contig of 4272 bp in length

12860: gap of unknown length

12860: contig of 4770 bp in length

17730: gap of unknown length

17830: contig of 5687 bp in length

23517: gap of unknown length

23616: contig of 3573 bp in length

27189: gap of unknown length

27289: contig of 3833 bp in length

31122: gap of unknown length

31222: contig of 4457 bp in length

31223: gap of unknown length

35679: contig of 3290 bp in length

35779: gap of unknown length

3780: gap of unknown length

39069: contig of 4254 bp in length

39169: gap of unknown length

39170: contig of 4398 bp in length

43424: gap of unknown length

43523: contig of 4398 bp in length

43524: gap of unknown length

47921: contig of 3342 bp in length

48021: gap of unknown length

48022: contig of 3076 bp in length

51363: gap of unknown length

51364: contig of 2725 bp in length

51464: gap of unknown length

54539: contig of 2725 bp in length

54540: gap of unknown length

54540: contig of 2725 bp in length

```
* 57365 57464: gap of unknown length
* 57465 61911: contig of 4447 bp in length
* 61912 62011: gap of unknown length
* 62011 66886: contig of 4475 bp in length
* 66886 66977: gap of unknown length
* 66977 70077: contig of 3391 bp in length
* 70077 73099: gap of unknown length
* 73099 73199: contig of 3022 bp in length
* 73199 76223: gap of unknown length
* 76223 76324: contig of 3024 bp in length
* 76324 78364: gap of unknown length
* 78364 78465: contig of 2041 bp in length
* 78465 80632: gap of unknown length
* 80632 80732: contig of 2168 bp in length
* 80732 83076: gap of unknown length
* 83076 83176: contig of 2344 bp in length
* 83176 86186: gap of unknown length
* 86186 86286: contig of 3010 bp in length
* 86286 89099: gap of unknown length
* 89099 91514: contig of 2813 bp in length
* 91514 91614: contig of 2315 bp in length
* 91614 93490: gap of unknown length
* 93490 93590: contig of 1876 bp in length
* 93590 95001: gap of unknown length
* 95001 95101: contig of 1411 bp in length
* 95101 97069: gap of unknown length
* 97069 97169: contig of 1968 bp in length
* 97169 99222: gap of unknown length
* 99222 102096: contig of 2053 bp in length
* 102096 102196: gap of unknown length
* 102196 103303: contig of 2774 bp in length
* 103303 103403: gap of unknown length
* 103403 105586: contig of 1107 bp in length
* 105586 105686: gap of unknown length
* 105686 107855: contig of 2183 bp in length
* 107855 107955: gap of unknown length
* 107955 110106: contig of 2151 bp in length
* 110106 112025: gap of unknown length
* 112025 112125: contig of 1819 bp in length
* 112125 113623: gap of unknown length
* 113623 113723: contig of 1498 bp in length
* 113723 115968: gap of unknown length
* 115968 116068: contig of 2245 bp in length
* 116068 118090: gap of unknown length
* 118090 118150: contig of 2022 bp in length
* 118150 119318: gap of unknown length
* 119318 119418: contig of 1128 bp in length
* 119418 120531: gap of unknown length
* 120531 120631: contig of 1113 bp in length
* 120631 122162: gap of unknown length
* 122162 122262: contig of 1531 bp in length
* 122262 124304: gap of unknown length
* 124304 124404: contig of 2042 bp in length
* 124404 126716: gap of unknown length
* 126716 126816: contig of 2312 bp in length
* 126816 127986: gap of unknown length
* 127986 128086: contig of 1170 bp in length
* 128086 129847: gap of unknown length
* 129847 129947: contig of 1761 bp in length
* 129947 131590: gap of unknown length
* 131590 131690: contig of 1643 bp in length
* 131690 133025: gap of unknown length
* 133025 133125: contig of 1335 bp in length
* 133125 134625: gap of unknown length
* 134625 134725: contig of 1500 bp in length
* 134725 135806: gap of unknown length
* 135806 135906: contig of 1081 bp in length
* 135906 137024: gap of unknown length
* 137024 137124: contig of 1118 bp in length
* 137124 137025: gap of unknown length
```

```
* 137125 138379: contig of 1255 bp in length
* 138380 138479: gap of unknown length
* 138480 139623: contig of 1144 bp in length
* 139624 139723: gap of unknown length
* 139724 140842: contig of 1119 bp in length
* 140843 140942: gap of unknown length
* 140943 142718: contig of 1776 bp in length
* 142719 142818: gap of unknown length
* 142819 143980: contig of 1162 bp in length
* 143981 144080: gap of unknown length
* 144081 145323: contig of 1243 bp in length
* 145324 145423: gap of unknown length
* 145424 146688: contig of 1285 bp in length
* 146689 146788: gap of unknown length

alignment_scores:
  Quality: 36.00      Length: 36
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-647-377-9 x AC098189  ..
  Align seg 1/1 to: AC098189 from: 1 to: 153701

      683 GlnPheargHisTyrrAlaLeuasnValProLeutyrrThrHisPheThrSe 699
      |||||
132037 CAGTTCGGGATTATGCTCTACACGTTCCTCTACACACACACACACACCTC 132086
      |||||

      699 rProileArgPheAlaAspValrleValHisArgLeuAlaAlaA 716
      |||||
132087 TCCCATCGCGCGCTTTGCTGAGCTCATAGTCACCGCTCCTCGCTGCTG 132136
      |||||

      716 IaLeuGly 718
      |||||
132137 CTCCTGGGT 132144

seq_name: gb_pr.AC019130
seq_documentation_block:
  LOCUS AC019130 139049 bp DNA linear PRI 10-JAN-2002
  DEFINITION Homo sapiens BAC clone RP11-563E16 from 2, complete sequence.
  ACCESSION AC019130
  VERSION AC019130.6 GI:18104914
  KEYWORDS HTG.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
  1 (bases 1 to 139049)
  AUTHORS Sulston,J.E. and Waterston,R.
  TITLE Toward a complete human genome sequence
  JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE 99063792
  REFERENCE
  2 (bases 1 to 139049)
  AUTHORS Levy,A. and Kozlowski,A.
  TITLE The sequence of Homo sapiens BAC clone RP11-563E16
  JOURNAL Unpublished (2002)
  REFERENCE
  3 (bases 1 to 139049)
  AUTHORS Waterston,R.H.
  TITLE Direct Submission
  JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
    MO 63108, USA
  REFERENCE
  4 (bases 1 to 139049)
  AUTHORS Waterston,R.
  TITLE Direct Submission
  JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington
    University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
    On Jan 10, 2002 this sequence version replaced gi:17432469.
  COMMENT
    ----- Genome Center
    Center: Washington University Genome Sequencing Center
    Center code: WUGSC
```

Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics -----  
 Center project name: H\_NH0563E16  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-1196K6; the clone sequenced to the right is RP11-95H14, 2000 bp overlap finished in RP11-563E16. Actual start of this clone is at base position 1 of RP11-563E16.

Unsure base at position 103812.

FEATURES  
 source  
 1. .139049  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="2"  
 /chromosome="2"  
 /clone="RP11-563E16"  
 /clone\_lib="RPCI-11"  
 1. .44

repeat\_region

1. .44  
 /rpt\_family="Alu"

repeat\_region

99. .119  
 /rpt\_family="(TTTA)n"

repeat\_region

101. .398  
 /rpt\_family="Alu"

repeat\_region

409. .704  
 /rpt\_family="L1"

repeat\_region

721. .771  
 /rpt\_family="L1"

repeat\_region

746. .925  
 /rpt\_family="L1"

repeat\_region

941. .1242  
 /rpt\_family="Alu"

repeat\_region

1391. .1465  
 /rpt\_family="Alu"

repeat\_region

1458. .1489  
 /rpt\_family="AT-rich"

repeat\_region

1469. .1766  
 /rpt\_family="Alu"

repeat\_region

1778. .1800  
 /rpt\_family="AT-rich"

repeat\_region

2123. .2203

repeat\_region  
 2276. .2578  
 /rpt\_family="Alu"  
 repeat\_region  
 2700. .2817  
 /rpt\_family="L1"  
 repeat\_region  
 3208. .3502  
 /rpt\_family="Alu"  
 repeat\_region  
 3719. .3917  
 /rpt\_family="MaLR"  
 repeat\_region  
 3941. .4132  
 /rpt\_family="MER1\_type"  
 repeat\_region  
 4133. .4435  
 /rpt\_family="Alu"  
 repeat\_region  
 4738. .4841  
 /rpt\_family="T-rich"  
 repeat\_region  
 4820. .4905  
 /rpt\_family="Alu"  
 repeat\_region  
 4906. .4928  
 /rpt\_family="(T)n"  
 repeat\_region  
 5216. .5443  
 /rpt\_family="Alu"  
 repeat\_region  
 5444. .6106  
 /rpt\_family="L1"  
 repeat\_region  
 6105. .6308  
 /rpt\_family="L1"  
 repeat\_region  
 6309. .6874  
 /rpt\_family="ERV1"  
 repeat\_region  
 6875. .7117  
 /rpt\_family="L1"  
 repeat\_region  
 7126. .7323  
 /rpt\_family="MIR"  
 repeat\_region  
 8351. .8380  
 /rpt\_family="(T)n"  
 repeat\_region  
 9846. .9910  
 /rpt\_family="ERV1"  
 repeat\_region  
 9884. .9912  
 /rpt\_family="(TG)n"  
 repeat\_region  
 9914. .10039  
 /rpt\_family="L1"  
 repeat\_region  
 10079. .10387  
 /rpt\_family="Alu"  
 repeat\_region  
 11512. .11717  
 /rpt\_family="Alu"  
 repeat\_region  
 11793. .11814  
 /rpt\_family="AT-rich"  
 repeat\_region  
 11802. .12092  
 /rpt\_family="Alu"  
 repeat\_region  
 12368. .12475  
 /rpt\_family="L2"  
 repeat\_region  
 13044. .13761  
 /rpt\_family="L2"  
 repeat\_region  
 13762. .14037  
 /rpt\_family="Alu"  
 repeat\_region  
 14038. .14056  
 /rpt\_family="L2"  
 repeat\_region  
 14414. .15007  
 /rpt\_family="L1"  
 repeat\_region  
 15012. .15189  
 /rpt\_family="L1"  
 repeat\_region  
 15302. .15354  
 /rpt\_family="T-rich"  
 repeat\_region  
 15331. .15589  
 /rpt\_family="Alu"  
 repeat\_region  
 15821. .16368  
 /rpt\_family="L1"  
 repeat\_region  
 16560. .16711  
 /rpt\_family="L2"  
 repeat\_region  
 17329. .17657  
 /rpt\_family="L2"  
 misc\_feature  
 17839. .18296  
 /note="match to EST AA651657 (NID:g2583309) ns89c12.sl"  
 repeat\_region  
 18178. .18292  
 /rpt\_family="L1"

```

repeat_region 18352..18638
               /rpt_family="L1"
repeat_region 18639..18991
               /rpt_family="MaLR"
repeat_region 18992..19535
               /rpt_family="L1"
repeat_region 19536..19745
               /rpt_family="MER1_type"
repeat_region 19746..20276
               /rpt_family="L1"
repeat_region 20330..20514
               /rpt_family="L1"
repeat_region 20518..20830
               /rpt_family="MER2_type"
repeat_region 20831..21130
               /rpt_family="L1"
repeat_region 21117..21231
               /rpt_family="(TTTCC)n"
repeat_region 21233..21269
               /rpt_family="(T)n"
repeat_region 21255..21389

alignment_scores:
  Quality: 30.00      Length: 30
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AC019130 ..

Align seg 1/1 to: AC019130 from: 1 to: 139049
      89 AspGlyAspArgAspIleAspGlyValValAlaArgAsnArgAl 105
      |||||||||||||||||||||||||||||||||||||||||||
50503 GATGGTGATCGAGACATTTTATTGATGGGGTTGTTGCTCGTAATAGAGC 50552

      105 aLeuAsnGlyAspLeuValValValLysLeuLeuProGlu 118
      |||||||||||||||||||||||||||||||||||||||||||
50553 CTTAAATGGGGATCTGGTGGTCTGTAAGAACTGCTTCCCGAG 50592

```



OM of: US-09-647-377-9 to: GenEmbl:\* out\_format : pfs  
Date: Jul 8, 2002 12:51 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODE=frame+2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spool/US09647377/runat\_08072002\_075944\_22292/app\_query.fasta\_1.940  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000  
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pcr  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09647377\_@CGN1\_112153 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-647-377-9  
Query length: 870  
Database: GenEmbl:\*  
Database sequences: 1797656  
Database length: 187333701  
Search time (sec): 2781.750000

score_list:	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat:AX015905	+ 4521.00	6459.68	0.0	3100	! AX015905 Sequence 8 from Patent
gb_pat:AX015910	+ 2185.50	3116.46	3.1e-165	1784	! AX015910 Sequence 13 from Patent
gb_pat:AX015899	+ 1944.50	2772.38	4.6e-146	1550	! AX015899 Sequence 1 from Patent
gb_in:U23514	+ 1500.00	2101.09	1.1e-108	35325	! U23514 Caenorhabditis elegans
gb_pl:SPAC2C4	+ 1198.50	1668.83	1.4e-84	35000	! 2992591 S.pombe chromosome I
gb_pat:AX015901	+ 1083.00	1540.29	2.0e-77	1140	! AX015901 Sequence 3 from Patent
gb_pl:AY057479	+ 1069.50	1510.48	8.9e-76	2992	! AY057479 Arabidopsis thaliana
gb_in:AY058669	+ 1051.50	1482.67	3.2e-74	3597	! AY058669 Drosophila melanogaster
gb_pl:AC012193	+ 1042.00	1435.12	1.4e-71	82454	! AC012193 Arabidopsis thaliana
gb_pr:AF330044	+ 1000.50	1401.84	1.0e-69	7320	! AF330044 Homo sapiens KIAA1008
gb_in:AY052150	+ 998.00	1407.46	4.9e-70	3129	! AY052150 Drosophila melanogaster
gb_pr:AB023225	+ 997.50	1401.18	1.1e-69	5231	! AB023225 Homo sapiens mRNA for
gb_htg:AC017755	+ 972.50	1330.37	9.6e-66	131853	! AC017755 Drosophila melanogaster
gb_in:AC093121	+ 972.50	1327.73	1.4e-65	168250	! AC093121 Drosophila melanogaster
gb_in:AC093501	+ 972.50	1327.40	1.4e-65	173494	! AC093501 Drosophila melanogaster
gb_in:AC105293	+ 972.50	1326.40	1.6e-65	190351	! AC105293 Drosophila melanogaster
gb_in:AE003467	+ 972.50	1321.52	3.0e-65	298640	! AE003467 Drosophila melanogaster
gb_in:DME276897	+ 960.00	1351.44	6.5e-67	3602	! A276897 Drosophila melanogaster
gb_htg:AC006169	+ 960.00	1309.43	1.4e-64	174279	! AC006169 Drosophila melanogaster
gb_pr:AK025634	+ 953.00	1346.15	1.3e-66	2324	! AK025634 Homo sapiens cDNA: FL
gb_pl:NCB18F11	+ 932.50	1277.24	8.8e-63	89222	! AL670011 Neurospora crassa DN
gb_pl:NCB10H4	+ 932.50	1277.06	9.0e-63	90763	! AL670010 Neurospora crassa DN
gb_pr:AB066536	+ 922.00	1302.34	3.5e-64	2189	! AB066536 Macaca fascicularis b
gb_htg:AC020006	+ 920.50	1253.92	1.7e-61	156905	! AC020006 Drosophila melanogaster
gb_in:AE008204	+ 920.50	1251.35	2.4e-61	199016	! AC008204 Drosophila melanogaster
gb_pl:DB76430	+ 919.00	1294.13	2.5e-61	204442	! AE003747 Drosophila melanogaster
gb_pl:SCYOL021C	+ 919.00	1289.83	1.0e-63	3140	! D76430 yeast DNA for DIS prot
gb_pr:HSM800675	+ 919.00	1289.83	1.7e-63	4673	! 274763 S.cerevisiae chromosome
gb_pl:YSPDI3P	+ 902.50	1274.19	1.3e-62	2228	! AL080158 Homo sapiens mRNA: ch
gb_pl:SPBC26H8	+ 902.50	1269.77	2.3e-62	3350	! M74034 Schizosaccharomyces pombe
gb_in:TB308998	+ 900.50	1243.04	7.0e-61	39520	! AC011743 S.pombe chromosome I
gb_pr:BC022089	+ 900.50	1266.92	3.3e-62	3344	! AJ308998 Trypanosoma brucei R
gb_htg:AC007577	+ 895.50	1219.89	4.1e-62	3694	! BC022089 Homo sapiens, Similar
gb_in:TB309002	+ 890.00	1136.02	6.4e-55	132635	! AC007577 Drosophila melanogaster
gb_in:U000057	+ 780.00	1090.59	2.2e-52	4630	! U00057 Caenorhabditis elegans
gb_pl:AF012898	+ 767.50	1072.29	2.3e-51	4794	! AF012898 Candida albicans prot
gb_htg:PFMAL13P2_2	+ 749.50	1012.55	4.9e-48	110000	! Continuation (3 of 4) of PFM
gb_in:CNS07EGA	+ 739.00	991.32	7.4e-47	194439	! AL590443 chromosome III of s
gb_pr:BC014124	+ 736.50	1035.04	2.7e-49	2466	! BC014124 Homo sapiens, clone I

gb\_pl:YSCSRK1A + 717.00 1001.32 2.1e-47 4196 ! M63004 S.cerevisiae SRK1 ge  
gb\_pat:AR013983 + 717.00 997.39 3.4e-47 6030 ! AR013983 Sequence 8 from pa  
gb\_pl:YSCSIT4B + 717.00 997.39 3.4e-47 6030 ! M60318 S.cerevisiae SSD1 pr  
gb\_pl:YSCD9819 - 717.00 982.01 2.4e-46 24942 ! U51031 Saccharomyces cerev  
gb\_bat:AP001519 - 626.50 825.16 1.3e-37 303650 ! AP001519 Bacillus halodur

seq\_name: gb\_pat:AX015905

seq\_documentation\_block:  
LOCUS AX015905 3100 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 8 from Patent WO9950284.  
ACCESSION AX015905  
VERSION AX015905.1 GI:10041648  
KEYWORDS house mouse.  
SOURCE Mus musculus.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 3100)  
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.  
TITLE Nucleic acid molecules which code proteins influencing bone development

JOURNAL Patent: WO 9950284-A 8 07-OCT-1999;  
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP ANDREAS (DE); WIRTH THOMAS (DE)

FEATURES  
source Location/Qualifiers  
1..3100  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
125..2737  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07620.1"  
/db\_xref="GI:10041649"  
/translation="MHPDYKLNLRSPGTRGVSSVGPNSVGPASGPGDKKKNKSMRG  
KRSIFETYSKEDVSEGLKRGTLQGVLRINPKAFHEAFIPSPGDRDIFIDGVVAR  
NKLNGDLVVVKLLPEDQWAKVKNESNDEATEADIPPEGCGHHPQQRKGSWG  
PDVIEAFDSDSEDRHGNTSLGVGVKLSISTDPGRKEOSTFVKNKDNTPIQD  
TRGLSEKSLAKSAVYIIEKKHSAATGILKLDKNSDLFKYALFSPSDHVRPFI  
VVLKDCPDQFMTRPKDFANTLFCIRIIMDKEDCNFALGQAKSLQSGAIEPTEGI  
LTVKGVDFSPFSEVLECPQLPMTIPDEVGKRRDLKDCIFTIDPSTARLDLAL  
ACRLRTDGTVEGVHADIYSYFVPGSSLDKVAERATSVLYLVOKVPMPLRLLCEEL  
CSLNPMTDKLTSFVINKLTPGCKILLEMFGRITIRSCTKSYDHAQSMIENTEKPE  
EELPSPESHSVEVHQAVNLNLSIAKOLRQRVDGALBLQDLKLAFTLDHETGLPO  
GCHIEYDVSSAGALKNSKLITGDDKYSILARKEVLTNNYSRPMOMALIFCSMLQDCE  
QMRHYALNVLPLYTHFTSPIRRFADVIVHRLAAALGYSEQDVEPTLQKQADHCNR  
RMAKRVQELSIGLTFVLVKGSPLESEAMVGVNLQAFDVLVLRFGVKRYICNAL  
ALRSYFQKVGKKPELTLVMEPDDLEEEPTQOVITIFSLVDVVLQAEATALKYSAILK  
RPGLEKASDEEPEP"

BASE COUNT 783 a 795 c 812 g 710 t  
ORIGIN

alignment\_scores:  
Quality: 4521.00 Length: 870  
Ratio: 5.197 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-647-377-9 x AX015905 ..

Align seg 1/1 to: AX015905 from: 1 to: 3100

1 MetAsnHisProAspTyrLysLeuAsnLeuArgSerProGlyThrProAr 17  
|||||  
125 ATGACACCTCTGACTACAGCTGACCTTCGGTCTCGGGGACCCCGAG 174

17 gGlyValSerSerValValGlyProSerAlaValGlyAlaSerProGlyA 34  
|||||  
175 AGGTGTGTCCTCTGCTGGTGGCCGAGTCTGCTGGTTCGCGGAGGTG 224

34 sPlsLysSerLysAsnLysSerMetArgGlyLysLysLysSerIlePhe 50

|||||  
225 ACAAAGTCAAAGAAAGTCCATCGGAGGGAAGAAAGAGCATATT 274  
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67  
|||||  
275 GAAACTCATGTCCAGGAGGATGTTTCAGAAAGCTTGAAGAGAGGAAC 324  
67 rLeuLeuGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlaP 84  
|||||  
325 ACTTATCCAGGGTGATTGAGAAATCAACCCAAAGAAAGTTTCATGAAGCCT 374  
84 heIleProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100  
|||||  
375 TCATTCTCTCCGGATGGTATCGGGACATTTTATTGATGGAGTTGTT 424  
101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuPr 117  
|||||  
425 GCTCGTAATAGAGCCTTAATATGGGACCTTGTGGTTGTAAACTGCTTCC 474  
117 oGluAspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIleG 134  
|||||  
475 TGAGGATCAGTGGAGGCGAGTCAACACGAGAGCAATGACAAAGAAATAG 524  
134 luAlaThrTyrGluAlaAspIleProGluGluGlyCysGlyHisPro 150  
|||||  
525 AAGCTTACTTATGAAGCTGACATCCCTGGAAGAGGGCTGTGGACATCACCCC 574  
151 LeuGlnGlnSerArgLysGlyTrpSerGlyProAspValIleIleGluAl 167  
|||||  
575 CTGCAGCAGTCCCGGAAGGCTGGAGTGGTCTGATGTCAATTATAGAGGC 624  
167 acIlnPheAspAspSerAspSerGluAspArgHisGlyAsnThrSerGlyL 184  
|||||  
625 TCAGTTTGTATGACAGCGACTCAGAAGATAGACATGGCAACACCATGGCC 674  
184 euValAspGlyValLysLysLeuSerIleSerThrProAspArgGlyLys 200  
|||||  
675 TGGTTGATGGTGTAAAGAAATGTCAATCTCTACTCTGCACAGGAAARA 724  
201 GluAspSerSerThrProValMetLysAspGluAsnThrProIleProG 217  
|||||  
725 GAAATTCATAGTACTCCAGTTATGAAGATGAGAACACCCCATACCA 774  
217 nAspThrArgGlyLeuSerGluLysSerLeuGlnLysSerAlaLysValV 234  
|||||  
775 GGACACAAGAGGCTTATCAGAAAGTCACTTCAGAAATCAGCAAGAGTGG 824  
234 alTyrIleLeuGluLysLysHisSerArgAlaAlaThrGlyIleLeuLys 250  
|||||  
825 TTTACATCTGGAGAAAGCAATCTCGAGCAGCAACTGGCATCTCGAAA 874  
251 LeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaLeuPheSe 267  
|||||  
875 CTCVTGGCTGATAAGAACAGTACCTGTTTAAAGAAATACGCCCTGTTC 924  
267 rProSerAspHisArgValProAtrIleTyrValProLeuLysAspCysP 284  
|||||  
925 TCCTTCAGACCACCGATGACCTAGAAATTTAGTACCTCTCAAGGACTGTC 974  
284 roGlnAspPheMetThrArgProLysAspPheAlaAsnThrLeuPheIle 300  
|||||  
975 CCCAGGACTTCATGACCCGACCTAAAGACTTTGCCAACACGCTGTTCATC 1024  
1025 TGCCGCATCATAGACTGAGAGGAGGACTGTAATTTTGCCTTGGGGCAACT 1074  
317 uAlaLysSerLeuGlyGlnAlaGlyGluIleGluProGluThrGluGlyI 334  
|||||  
1075 GGCTAAGAGTCTTGGCGAGGCTGGTGAATCGAGCCTGAAACAGAGGGA 1124  
334 leLeuThrGluTyrGlyValAspPheSerAspPheSerGluValLeu 350  
|||||

1125 TACTGACAGAAATATGGTGGGACTTCTCTGATTTCTCTTCAGAAAGTTCTT 1174  
351 GluCysLeuProGlnSerLeuProTrpThrIleProProAspGluValG 367  
|||||  
1175 GAATGTCTCCCTCAAAGCTGCCCTGGACAAATCCACCTGATGAGGTGGG 1224  
367 yLysArgArgAspLeuArgLysAspCysIlePheThrIleAspProSerT 384  
|||||  
1225 CAAGAGAAGAGACCTAAGAAAGACTGTATCTTCACCAATTGATCCATCAA 1274  
384 hrAlaArgAspLeuAspAspAlaLeuAlaCysArgArgLeuThrAspGly 400  
|||||  
1275 CTGCTCGCACCTTGTATGATGCCCTCGCTGCAGCGGCTCACTGATGGC 1324  
401 ThrPheGluValGlyValHisIleAlaAspValSerTyrPheValProG 417  
|||||  
1325 ACCTTCGAAGTGGGGCTCCACATCGCCGATGTGAGTTACTTTGTTCCTGA 1374  
1375 GGGATCCTCTTTGGATAAAGTAGCTGTGAGAGAGCCACAAGTGTCTACT 1424  
434 euValGlnLysValValProMetLeuProArgLeuLeuCysGluGluLeu 450  
|||||  
1425 TGCTCCAGAAGGTGTCCCATGCTTCCAGAGGCTCTGTGTGAGAACTC 1474  
451 CysSerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLy 467  
|||||  
1475 TGAGGCTCAACCCCATGACTGACAAGCTGACCTTCTCTGTGATCTGGAA 1524  
467 sLeuThrProGluGlyLysIleLeuGluGluTrpPheGlyArgThrIleI 484  
|||||  
1525 GCTGACCCCTGAAGCAAGATCCTTGAAGAGTGGTTTGGCCGCACTATCA 1574  
484 leArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGlu 500  
|||||  
1575 TCGTTCITGACCAAACTGAGCTAGCACCATGCCAGAGCATGATCGAA 1624  
501 AsnProThrGluLysIleProGluGluGluLeuProProfileSerProG 517  
|||||  
1625 AATCCAACGTGAAGATCCCTGAGGAAGAGCTTCCCCCAATTTCTCCAGA 1674  
517 uHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleA 534  
|||||  
1675 GCACAGCTTCAGAGAGTGCACCGGAGTCTCTGAACCTGCACAGCATG 1724  
534 laLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAsp 550  
|||||  
1725 CAAAGCAACTCGCCGCCAGCGCTTTGTAGATGGGCACCTCCGTTTAGAT 1774  
551 GlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnG 567  
|||||  
1775 CAGCTGAAGCTGCTTTTACTCTGGACCATGAGACTGGACTGCCCTCAAG 1824  
567 yCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPheM 584  
|||||  
1825 ATGTACATCTATGAGTACCGAGACAGCAACAAGCTTGTAGAGAGTTCA 1874  
584 etLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhePro 600  
|||||  
1875 TGCTCTCGCCCAACATGGCGGTGGCCCAAGATCTTCCGACCTTCCCT 1924  
601 GluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSe 617  
|||||  
1925 GAGCAGGCCCTGTGCGCGGCAATCCCCACACAGAGCAAGATGTCTCAG 1974  
617 rAspLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSerS 634  
|||||  
1975 TGACCTGGTGGAGTTCTGTGACCAGATGGGCTGCCCATGATGTCAAGT 2024  
634 erAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspLys 650  
|||||  
2025 CTGCAGGGGGCCTAAATAAAGCCTGACTAAGACATTTGGAGATGACAAG 2074

651 TyrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMe 667  
|||||  
2075 TACTCTCTGGCCGGAAGGAGGTGCTCACCAACATGTACTCCCGCCCAT 2124  
|||||  
667 tGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGlnP 684  
|||||  
2125 GCAGATGGCACTGTACTCTGCTCTGGATGCTGAGGACGAGGAGCT 2174  
|||||  
684 heArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerPro 700  
|||||  
2175 TCCGGCATATGCTCTCAAGTTCCTCTACACACATTCACCTCTCCC 2224  
|||||  
701 ileArgGlyPheAlaAspValIleValHisArgLeuLeuAlaAlaLe 717  
|||||  
2225 ATCCGCGCTTGTGCTGACGTCAATGTCACCGCTCTCTGGCTGCTCT 2274  
|||||  
717 uGlyTyrSerGluGlnProAspValGluProAspThrLeuGlnLysGlnA 734  
|||||  
2275 GGGCTACAGTGAACAGCCAGATGTGAGGCTGTATACCTACAGAGCAAG 2324  
|||||  
734 laAspHisCysAsnAspArgArgMetAlaSerLysArgValGlnGlnLeu 750  
|||||  
2325 CTGACCACTGCAATGACGCTGCATGCTTCCAAAGTGTGCAGAGCTC 2374  
|||||  
751 SerIleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuGl 767  
|||||  
2375 AGCATCGGCTCTCTTCGCGAGTTCTAGTAAAGGAGAGTGGCCCCCTGGA 2424  
|||||  
767 uSerGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuV 784  
|||||  
2425 GTCCGAGGCATGTTGATGGGTGCTGAACCAAGCTTTCGAGCTGCTGG 2474  
|||||  
784 allLeuArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeu 800  
|||||  
2475 TGTGCGGCTTTGGGTGAGAAGGCGCATCTACTGCAATGCACCTGGCCCTG 2524  
|||||  
801 ArgSerTyrSerPheGlnLysValGlyLysProGluLeuThrLeuVa 817  
|||||  
2525 CGATCTACAGCTTCCAGAGGTGGGAGAAGCCAGAGCTCCTCTTGT 2574  
|||||  
817 lTrpGluProAspAspLeuGluGluProThrGlnGlnValIleThrI 834  
|||||  
2575 TTGGAGCCTGATGACCTTGAAGAGAGGACCAACACAGAGGTATCACCA 2624  
|||||  
834 lePheSerLeuValAspValLeuGlnAlaGluAlaThrAlaLeuLys 850  
|||||  
2625 TCTTCAGCCTGGTGGATGTGCTCTGCAGGCAGAGGCCACAGCCCTCAAG 2674  
|||||  
851 TyrSerAlaIleLeuLysArgProGlyLeuGluLysAlaSerAspGluGl 867  
|||||  
2675 TACAGTGCTATCTGAAGCGACAGGCTTGAGAAAGGCTGCTGATGAGGA 2724  
|||||  
867 uProGluAsp 870  
|||||  
2725 GCCTGAGGAC 2734

seq\_name: gb\_pat:AX015910

seq\_documentation\_block:  
LOCUS AX015910 1784 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 13 from Patent WO9950284.  
ACCESSION AX015910  
VERSION AX015910.1 GI:10041653  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1784)  
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.  
TITLE Nucleic acid molecules which code proteins influencing bone development

JOURNAL Patent: WO 9950284-A 13 07-OCT-1999;  
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP  
ANDREAS (DE); WIRTH THOMAS (DE)  
FEATURES  
source  
1..1784  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<3..1454  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07621.1"  
/db\_xref="GI:10041654"  
/translation="NFKGVHIADVSVFVPEGSLDKVAERATSYLVVQKVPMLPR  
LLCEELSLNPMDSKLTFSVITWITPPEKILDEWFGRTIIRSCTKLSYEHQAQSMIESP  
TEKIPAKELPPI SPEHSSEVHQAVNLHGIKOLRQRFVDGALLRDLQKLAFTLDH  
ETGLPOGCHIEYRESNKLVEEFMILLANNAVAHKIHRAPFEQALLRRHPPTQRMLS  
LVFCDOMGLPVDFSAGALSKLTQTGGDKYSLARKEVLTNMSRPMQALYFCSG  
LLQDPAGFRHYALNVPLYTHFTSPIRREADVLRHLAAALGYRRFLDMAPDTLQKOA  
DHCDNRMAKSRVQELSTSLFFAVLYKESGPLESEAMVGLIKQAFDVLVLYGVOKR  
IYNALAIRSHRFQVKRPKPELTLVWPEFMEQEPAAQOQVITFSLVEVVLQNESTALK  
YSAILKRPGTQGHLPKEEESDGEPEDSST"  
BASE COUNT 385 a 537 c 504 g 358 t  
ORIGIN

alignment\_scores:  
Quality: 2185.50 Length: 478  
Ratio: 4.825 Gaps: 1  
Percent Similarity: 94.770 Percent identity: 88.494

alignment\_block:  
US-09-647-377-9 x AX015910 ..  
Align seg 1/1 to: AX015910 from: 1 to: 1784  
402 PheGluValGlyValHisIleAlaAspValSerTyrPheValProGluGl 418  
|||||  
6 TTCAAAGTGGGAGTTCACATTGCTGACGTGAGTTACTTTGTTCCGGAGGG 55  
|||||  
418 ySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrLeuV 435  
|||||  
56 ATCTGATCTGGATPAAAGTGGCTGCCGAGAGGCTACAAAGCGTCTACTTGG 105  
|||||  
435 alGlnLysValValProMetLeuProArgLeuLeuCysGluGluLeuCys 451  
|||||  
106 TTCAAAGGTGGTCCCATGCTTCCAGCGCTGCTGTGTGAGAGCTGTGC 155  
|||||  
452 SerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLysLe 468  
|||||  
156 AGCCTCAACCCCATGTCCGACAGCTGACCTTCTCTGTGATCTGGACACT 205  
|||||  
468 uThrProGluGlyLysIleLeuGluTrpPheGlyArgThrIleIleA 485  
|||||  
206 GACTCCAGAGGGCAAGATCCTTGATGAATGGTTGGCGCGACCATCATCC 255  
|||||  
485 rgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGluAsn 501  
|||||  
256 GCTCTCTGCCCAAACTTAGCTACGACATGCACAGCATGATTGAAAGC 305  
|||||  
502 ProThrGluLysIleProGluGluGluLeuProProIleSerProGluHi 518  
|||||  
306 CCAACTGAGAAAATCCCTGCGAAAGAGCTGCCCCCATTTCCCCAGAGCA 355  
|||||  
518 sSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleAlaL 535  
|||||  
356 TAGCAGCGGAGGGGTACACCAGCGCGCTTGAATCTCCACGGGAATTGCCA 405  
|||||  
535 ysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAspGln 551  
|||||  
406 AGCACTTACGCCACAGCGCTTTGTGGACGGCGCACTTCGTTTGGATCAG 455  
|||||  
552 LeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnGlyCy 568  
|||||

```
456 CTAAGCTTGCTTCACTCTGGACCACGAGACCGGATTGCTCAAGGATG 505
568 sHsIleTyRgLUtyrArgAspSerAsnLysLeuValGluGluPheMetL 585
||||:|||||
506 TCATATCTATGAGTACCGCAGAGCAACAGCTCTGGAGAGTTCATGC 555
585 euLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPheProGlu 601
556 TCTTGCCCAACATGGCAGTGGCCACAAAGATCCACCGCGCTTCCCGAG 605
602 GlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSerAs 618
|||||
606 CAGGCCCTGCTGCGCCGCGCACCCCGCCCAACAGGATGCTCAGTGA 655
618 pLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSerA 635
656 CCTGGTGAATTCCTGCACACAGATGGGCTGCCGTGACTTCAGTCCG 705
635 laGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspLysTy 651
|||||
706 CAGGAGCCCTCAATAAAGCCTGACCCAAACATTTGGAGATGACAAGTAC 755
652 SerLeuAlaArgLysGluValLeuThrAsnMetTyRserArgProMetG 668
756 TCACCTGGCCCGCAAGGAGGTGCTCAACCAACATGTCTCCGGGCCATGCA 805
668 nMetAlaLeuTyRpheCysSerGlyMetLeuGlnAspGlnGluPheA 685
806 GATGGCATGTACTTCTGCTCGGGCTGCTGCAGGACCCAGCGCAGTTCC 855
685 rgHisTyRAlaLeuAsnValProLeuTyRThrHisPheThrSerProIle 701
856 GCACATAGCGGCTCAATGTGCCCTGTATACACACATTCACCTGCCCATC 905
702 ArgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuG 718
|||||
906 CGCCGCTTTCGCGAGTCTCTGTGTGACCGCCTCTCTGGCTGCGCGTTAG 955
718 yTyRserGluGlnProAspValGluProAspThrLeuGlnLysGlnAla 735
956 CTATAGGAGCGGCTAGATGATGCCGCCGATACCCCTGCAGAAACAGCGG 1005
735 sPhisCysAsnAspArgArgMetAlaSerLysArgValGlnGluLeuSer 751
1006 ACCACTGTAAAGCCCGCCGATGGCTCCAAAGCGGTGCAGGAGCTCAGT 1055
752 IleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuGluSe 768
1056 ACCAGTCTCTCTTGTGTGTGTCAAGGAGAGTGGCCCTGGAGTC 1105
768 rGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuVal 785
1106 AGAAGCCATGTATGGCATCCTAGACAAAGCCTTCGACGTGCTGTCG 1155
785 euArgPheGlyValGlnLysArgIleTyRcysAsnAlaLeuAlaLeuArg 801
1156 TGCGCTACGGCTGCAGAGACGCATCTACTCAACGCACTGGCCCTGCGG 1205
802 SerTyRserPheGlnLysValGlyLysLysProGluLeuThrLeuVal 818
1206 TCCCAACCACTTCCAGAGGTGGCAAGAGCCGGAACCTCAGCTGCTGTG 1255
818 pGluProAspLeuGluGluProThrGlnGlnValIleThrIleP 835
1256 GGAGCCTGAGGACATGGACAGGAGCCACAGCAGGATCATCACCATCT 1305
835 heSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLysTy 851
1306 TCAGCCTGGTGAGGTGGTCTGCAGGAGAGTCCACAGCCCTCAAGTAC 1355
852 SerAlaIleLeuLysArgProGlyLeu..... 860
1356 AGCGCCATCTGAAGCGGCCAGGCACCCAGGGCCACCTGGGCCCTGAGAA 1405
```

```
861 ....GluLysAlaSerAspGluGluProGluAsp 870
||||:|||||
1406 GGAGGAGGAGGAGTCTGACGGTGGCCCGAGGAC 1439

seq_name: gb_pat:AX015899

seq_documentation_block:
LOCUS AX015899 1550 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from Patent WO9950284.
ACCESSION AX015899
VERSION AX015899.1 GI:10041641
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1550)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL PATENT: WO 9950284-A 1 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
source
1..1550
/organism="Mus musculus"
/db_xref="taxon:10090"
<2..1183
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC07618.1"
/db_xref="GI:10041642"
/translation="LGRSKVAABRATSYLVQKVVPMPLRLLCEELCSLNPMDTKLTF
SVIWKLTPEGKILEWFGRTIIRSCTNLSDHAQSMIENTEKIPEELPPTISPEHSV
EEVHQAVLNLSIAKQLRQRFDVGLRDLDDQEFMLLANVAHKIFRTFPEQALLRRH
PPQTKMLSDLVFECQDQGLPMDVSSAGALNNMALYFCSMQLQDQEQFRHYALNPLYT
HFTSPILRRFADYIVHRLIAALGYSEQDPVEDPTLQKQACHNDRMRASRVQELSIG
LFFAVLVKESGPLEAMVGLNQAFDLVLRFGVKRIYCNALALRSYSFQKVGKK
PELTLWEPDLEETQOVITIFSLVDVVVLAQENTALKYSAILLRPGLEKASDEEPE
D"
BASE COUNT 352 a 433 c 412 g 353 t
ORIGIN

alignment_scores:
Quality: 1944.50 Length: 448
Ratio: 4.999 Gaps: 2
Percent Similarity: 86.830 Percent Identity: 86.830

alignment_block:
US-09-647-377-9 x AX015899 ..
Align seg 1/1 to: AX015899 from: 1 to: 1550
423 LysValAlaAlaGluArgAlaThrSerValTyRLeuValGlnLysValVa 439
|||||
14 AAAGTAGCTGCTGAGAGGCCACAAAGTGTCTACTTGGTCCAGAAAGTGGT 63
439 lProMetLeuProArgLeuLeuCysGluGluLeuCysSerLeuAsnProm 456
64 CCCCATGTCTCCAGGCTTCTGTGTGAGGAACCTCGCAGCCTCAACCCCA 113
456 eThrAspLysLeuThrPheSerValIleTyRLeuThrProGluGly 472
114 TGACTGCACAGCTGACCTTCTCTGTGATCTGGAAGCTGACCCCTGAAGGC 163
473 LysIleLeuGluGluTyRPhGlyArgThrIleIleArgSerCysThrLy 489
164 AAAGTCCCTTGAAGAGTGGTTTGGCCGCACTCATCTCTTTCACCAA 213
489 sLeuSerTyRAspHisAlaGlnSerMetIleGluAsnProThrGluLysI 506
|||||
```

214 ACTGAGCTACGACCATGCCAGAGCATGATCGAAATCCAACTGAGAAGA 263  
506 leProGluGluLeuProGluLeuProGluLeuProGluLeuProGluLeu 522  
264 TCCTGAGGAGAGGCTCCCAATTTCTCCAGAGCAGCAGCTCGAGGAG 313  
523 ValHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArg 539  
314 GTCCACCGGCGCTCTCAACCTGCACAGCATTTGCAAGCAACTCCGCG 363  
539 gGlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaP 556  
364 CCAGCGCTTTGATGAGGCGCAGCTCCGTTTAGATCAG..... 400  
556. heThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrGlu 572  
400 ..... 400  
573 TyrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsnMe 589  
401 .....GAGTTTCATGCTCTCTGGCCCAACAT 423  
589 talValAlaHisLysIlePheArgThrPheProGluGlnAlaLeuLeuA 606  
424 GCGGCTGCCCAAGATCTCCGACCTTCCCTGAGCAGGCCCTGCTGC 473  
606 rgArgHisProProGlnThrLysMetLeuSerAspLeuValGluPhe 622  
474 GCGCGCATCCCCACACAGCAGAGATGCTCAGTGCCTGGTGGAGTTC 523  
623 CysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeuAs 639  
524 TGTGACCATGGGCTGCCATGGATGTCAGTCTGCAGGGGCCCTAAA 573  
639 nLysSerLeuThrLysThrPheGlyAspAspLysTyrSerLeuAlaArgL 656  
574 T..... 574  
656 ysGluValLeuThrAsnMetTyrSerArgProMetGlnMetAlaLeuTyr 672  
575 .....ATGGCACTGTAC 586  
673 PheCysSerGlyMetLeuGlnAspGlnGluGlnPheArgHisTyrAlaLe 689  
587 TTCTGCTCTGGATGTCGAGACAGCAGCAGTTCGGCATATGCTCT 636  
689 uAsnValProLeuTyrThrHisPheThrSerProIleArgA-gPheAlaA 706  
637 CAACGTTCCCTCTACACACACTTCACCTCTCCCATCCGCGCTTGCTG 686  
706 spValIleValHisArgLeuLeuAlaAlaLeuGlyTyrSerGluGln 722  
687 ACGTCATAGTCACCGCTCTCTGGCTGCTCTGGCTACAGTGAACAG 736  
723 ProAspValGluProAspThrLeuGlnLysGlnAlaAspHisCysAsnAs 739  
737 CCAGATGTGAGCGCTGATACCTACAGAACAGCAAGCTGACCACTGCAATGA 786  
739 pArgArgMetAlaSerLysArgValGlnGluLeuSerIleGlyLeuPheP 756  
787 CCGTCGATGGCTTCCTCAACAGGTGCGAGAGCTCAGCATCGGCTCTTCT 836  
756 heAlaValLeuValLysGluSerGlyProLeuGluSerGluAlaMetVal 772  
837 TCGCAGTTCTAGTAAGGAGAGTGGCCCTCTGGAGTCCGAAGCATGTGTG 886  
773 MetGlyValLeuAsnGlnAlaPheAspValLeuValLeuArgPheGlyVa 789  
887 ATGGGTGCTCTGAACCAAGCTTTCAGCTGCTGGTGGCTTTGGGT 936  
789 lGlnLysArgIleTyrCysAspAlaLeuAlaLeuArgSerTyrSerPheG 806  
937 GCAGAGCGCATCTACTGCAATGCACCTGGCCCTCGCATCTCAGCTTCC 986

806 InLysValGlyLysLysProGluLeuThrLeuValTyrGluProAspAsp 822  
987 AGAGGTGGGAGAGAGCCAGAGCTCCTGTTTGGAGCCTGATGAC 1036  
823 LeuGluGluGluProThrGlnGlnValIleThrIlePheSerLeuValAs 839  
1037 CTTGAGAGGAGCCACACAGCAGGTCTACCATCTTCAGCCTGGTGA 1086  
839 pValValLeuGlnAlaGluAlaThrAlaLeuLysTyrSerAlaIleLeuL 856  
1087 TGTGCTCTCGAGCGCAGAGCCACAGCCCTCAAGTACAGTCTATCTGA 1136  
856 ysArgProGlyLeuGluLysAlaSerAspGluGluProGluAsp 870  
1137 AGCGACCAAGCCTGGAGAGCGCTGATGAGGAGCCTGAGGAC 1180  
seq\_name: gb\_in:U23514  
seq\_documentation\_block:  
LOCUS U23514 35325 bp DNA linear INV 16-NOV-2001  
DEFINITION Caenorhabditis elegans cosmid F48E8, complete sequence.  
ACCESSION U23514  
VERSION U23514.2 GI:16950416  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE  
AUTHORS The C. elegans Sequencing Consortium.  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)  
JOURNAL 99069613  
MEDLINE  
AUTHORS 2 (bases 1 to 35325)  
TITLE The sequence of C. elegans cosmid F48E8  
JOURNAL Unpublished  
AUTHORS 3 (bases 1 to 35325)  
TITLE Waterston, R.  
AUTHORS Direct Submission  
JOURNAL Unpublished  
REFERENCE 4 (bases 1 to 35325)  
TITLE Waterston, R.  
AUTHORS Direct Submission  
JOURNAL Submitted (27-MAR-1995)  
REFERENCE 5 (bases 1 to 35325)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUL-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
REFERENCE 6 (bases 1 to 35325)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
REFERENCE 7 (bases 1 to 35325)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
REFERENCE 8 (bases 1 to 35325)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
COMMENT On Nov 16, 2001 this sequence version replaced gi:746484.  
Submitted by:

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rwnematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

#### NEIGHBORING COSMID INFORMATION

The clone sequenced to the left is T24G10, 1600 bp overlap; the  
clone sequenced to the right is R02F2, 200 bp overlap. Actual  
start of this clone is at base position 1597 of CELF48E8; actual  
end is at 3296 of CELR02F2.

#### NOTES:

Coding sequences below are predicted from computer analysis, using  
the program Genefinder (P. Green and L. Hillier, ms in preparation).  
Location/Qualifiers

```

1. 35325
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="III"
/clone="F48E8"
75..3061
/gene="F48E8.6"
join(75..395,648..1856,2043..2522,2568..2756,2807..3061)
/gene="F48E8.6"
/notes="similar to S. cerevisiae SSD1 protein
(SP:SSD1_YEAST, P24276) and to E. coli VACB and
Ribonuclease II genes: contains similarity to Pfam domain
PF00773 (RNB), Score=680.8, E-value=2.2e-201, N=1;
for by the following C. elegans cDNAs: yk14e11.3,
yk190a9.3, yk219h5.5, yk226g9.5, yk309h10.5, yk317f1.3,
yk317f1.5, yk347e5.5, yk358h7.3, yk358h7.5, yk395c2.5,
yk444d2.5, B1175755, yk545b2.3, yk545b2.5"
/codon_start=1
/product="Hypothetical protein F48E8.6"
/protein_id="AAC46543.1"
/db_xref="GI:746490"
/translation="MFRVKTPTTRAKPVKPNQANNLONNPNRKFITYEYSKEETD
AGEDSMFGVLRIKPNKYQECFLDHPKGNHPDVLVLGODRNRAMQGDVVAVKIP
KEDLVNVEYVKKWAEHKKGRNSKTDNNSPNKRCLRNEIQDNGVTSDEVPDS
CLITITGALVHLEKHPVAAAGKLOIMPNSANPNVLFVATDSRVPRLIPKSDVDKEF
FSRQDFRFLYTAKITDRAESVYADGRVLKLLGMSGETDTEFIVVEHOIDHREF
SDCELESIPITTAENKVPDPAFEYRDRPSDVIPTIDPRTARDLDDALHAKHIDCD
GKPTGLEIGHVADYFELEKGETELDKWASERGNSTYLSQTVIMPLRICEQLCSL
NPGVDLRSFIVFKMSYAEELIDVWFGSRVSKRSYKRLAYEHAQDFIENPEKDTCEDEL
PDIDGNTPFEIKRITLMURLHIAQVLRQKREDSGALRIELPRKALDEBKKPGQVSI
YEIDSKNLVEEFLMANVEAKKTAENPPEHALLRNHPPEKMKIDVAEQCARIGE
PLDGTSGLLSTLRKYGSKSLDMCIRQVSSLTIRPMQAKYCFCTFEMPFSFHHF
ALNVDYTHFTSPIRTPDPIVHRQALAAAGYNERSERVPEEQIEICTNCDTKLASK
EASDESAMLYFGVFIHQTPGMKCOAVLGVMDLFDVLIVEYGVKRVVYDKMKRDEN
KSTEKLITYPADPNAESGREFSSIQMCNVVYVILVPKSYEVSATIVRPSLEQR
NLLKSTLKKMETGSTILO"
complement(3404..5011)
/gene="F48E8.7"
complement(join(3404..3599,3650..3812,3863..3996,
4106..4392,4436..4760,4808..4889,4933..5011))
/gene="F48E8.7"

```

gene

CDS

FEATURES  
source

gene

CDS

```

/notes="contains similarity to Pfam domain PF00646 (F-box),
E-value=0.023, N=3; coded for by the following C. elegans
cDNAs: yk120d4.3, yk120d4.5, yk261a3.3, yk261a3.5,
yk366f7.3, yk366f7.5, yk423a12.3, yk423a12.5, yk497a12.3,
yk513f4.5, yk513f4.3"
/codon_start=1
/product="Hypothetical protein F48E8.7"
/protein_id="AAC46542.2"
/db_xref="GI:14670174"
/translation="MSRVIGRLPSTLEDNGPSENEHCVETVFSDEFTSIPPCYNLYPK
KSLRUSVSGNSKRLSNLNCNLVDEDEIAPIKRKDARSIVTFVSDFTSIPPCYNLYPK
SNKKDLDLSAMLVCHRYFGTGHKSRNWITDAQDRPISEISIALSORKIKLVRLAG
AKADRLADARLVEVALLSTSRLELLDRTNLRTARQIMLKPKQKLCISLEGNO
LDHVAITAENRGLRELDISWTNGANGSLIFRNCKDLQOLNLCWAGTQPIVTV
IIFISGEKLLKLNLSGTVRNGINNKHEILAAKSNVVTDLSDNVSEISDGLAVIL
NKPRLTITSLARCXGLDPNMVHFNSKPSLVYLVNHHGCVSTNNLEFLCNSLNKN
ROYENTAKPVSYQCESNMINGQTMMDIE"
6868..9530
/gene="F48E8.5"
join(6868..6983,7253..7347,7498..8017,8069..8928,
8976..9193,9247..9326,9371..9462,9517..9530)
/gene="F48E8.5"
/notes="coded for by the following C. elegans cDNAs:
yk11.5, yk12f11.3, yk12f11.5, yk37g2.3, yk37g2.5,
yk37h5.3, yk37h5.5, yk62c2.3, yk62c2.5, yk75a2.3,
yk75a2.5, yk85f11.3, yk85f11.5, yk97f8.3, yk97f8.5,
yk104a7.3, yk104a7.5, yk145c4.5, yk145c4.3, yk182a6.5,
yk288a6.3, yk288h7.3, yk288h7.5, yk286f3.5, yk286f3.5,
yk288d12.3, yk288d12.5, yk288e9.5, yk290h8.5, yk357d4.5,
yk374a3.5, yk429e12.3, yk429e12.5, yk430f1.3, yk430f1.5,
yk447a6.5, yk472d10.5, yk474b10.5, yk479c12.5, yk490c6.5,
yk491b2.5, yk502g6.3"
/codon_start=1
/product="Hypothetical protein F48E8.5"
/protein_id="AAC46541.1"
/db_xref="GI:746488"
/translation="MDCAIAYRWRPFESASLRITVAASVLLSLPLFCFVRYSKMT
IRRFQFFQNTQFFFLSNLQKLPALFTMSVEEATDDALYPIAVLIDELRNEVDV
LRNLRKSLTALAGVERTRNELIOFLDTIYDEDEVLVLAEOGLNFTPLVGPD
VHCLLLPLENATVEETVVRDKAVESLRKADKSSASLEHFPVRLRANGDMFT
SVTSAGLFSVYVPSYPAIKSELKSMFRLCDRDPWVRRAAAKLGEPKVEKTA
VIGLHSSUTLDRHVEDQSVRLTVESATFGLDKKANKKLELILIELFDKSWR
VRYVAAKLEIQNVLGDMDTLHVNNTNLLKDPGEVRCATQRLQEFALNLPED
KRONICNSLNLVAKELVDGNQLVKSAGVIMGLAPIKEQTVSELPIYXQLIN
DOPEVRLNIISLKVNEVGAALSTLLPAVLAEDGKRWVRVLAIVQFNPPLAS
QLQGEFFDLKLLPCLNLTLDHVSIREASTLIMKELTOKFGGOWASTNIVPKMKLO
KDTNYLQMTCLFLCNTLSSEAMTQEQILKEIMPVYKDLVEDDVPVNRNAAKSLKRIG
KNLTPTSLTSEVKKPLEKLGKSDFDVRYFSEAKNSLGL"
10374..15737
/gene="F48E8.4"
join(10374..10667,10934..11087,11140..11312,11753..12155,
12494..12600,12650..12735,13479..13739,13786..14074,
14454..14626,14736..14835,15148..15433,15598..15737)
/gene="F48E8.4"
/codon_start=1
/product="Hypothetical protein F48E8.4"
/protein_id="AAC46544.1"
/db_xref="GI:746491"
/translation="MIITFLOLFLPSTATATAGVDFLAEYRVSVDANIGENLWEPAP
ELDIRVTASIDDLLEKRVINRPDDSTONLMTFVTSKGQKFCASLPDVEDYKDKP
KSKNPKIYGDALAAFYVDRVKLRGNHWSYILRCGQTIQVHGPGGEGYVKNIL
GLFDGSKMPSYESTDRLLEVEETYSQTCFDEEYREPMTSVREYCEQAOLSTNE
VYTSVVEYKPCOYLMIVKVTGLCYPEFLPASQNTKKIGQCFLEKEDVROLLRQ
LEBKQAESQSVLTVKRRFORAERFANMTKSNRNLIKPDVVKRTNAMEDVHAHF
NLLVAGLADYDVRDENRGLWYFHDSPWPNRPNOKDMATATQNGYLSQVOKNLG
TQRFQNDDELVEGLTHFKTGELDLSNLVIPAMDIDEMWFIQVRQDPTKWLIDLY
EFTHEDRDLVAGSLRAQDYIEGMLHLKLPKTKSWFPMGNLRNSLVILDEVEDDI
IELSTNGLESLDSSSNYKAIYDWDIGNASPOLKTKNGKLNKFTVYSQARDY
EIDAEQQLLPTIYKLSVSMALIQWFCQRPNRPHVHPPEEHELEPSTDTEFIFHM
NLARDMSRFARLMTFVFKVQDQTRWRNDKPAERFVEDFTVWMPHATTDANEA
FINLHKDLGPOLILLEKERTWNSKVEKYRTYWKDYKMRQAOSMOGLMKDLKMD
KDAFVYSQSSATTTKQASPFDELLKMSSESDLSSELLESLSKNAAGVEEKVEPFDM

```

gene

CDS



```
1635 TTATCATGAGAACCCGGAGAGGAGCTTCACATCGCAGCAGCTTCCCGATAT 1684
514 eSerProGluHisSerValGluGluValHisGlnAlaValLeuAsnLeuH 531
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1685 TTCAGATGGAAATACGCCCTTTGAGATCAAGGAAAGACTCTGATGTTC 1734
531 isSerIleAlaLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeu 547
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : :
1735 ACAGAATCCGCAAGTTCTTCGACAAAAGCGTGAAGATAGTGGACATTG 1784
548 ArgLeuAspGlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLe 564
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1785 -CGATCGACTCCAGATGAAATTTGCATTGGAT...GAGATTAAGAA 1831
564 uProGlnGlyCysHisIleTyrGlu..... 572
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1832 GCCACAGGAGTTTCTATATACAGAGTTAATAAACAGTTTTTTTGTAAAG 1881
572 ..... 572
1882 AAAATGTTTATATTAATTTTGTTCACACAGCAATTTTGTGTC 1931
572 ..... 572
1932 ACATTTTTCGGTTCTAGCAACAAAAAGTAATCCATAAATTTAGATTTT 1981
572 ..... 572
1982 CAAAATCGGAATTTAAAGAAATTCGCAATTTTATAAAATACAAATGT 2031
573 .....TyrArgAspSerAsnLysLeuValGluGluPheMetLeu 585
: : : : : : : : : : : : : : : : : : : : : : : : : :
2032 TAAATTTTCAGATCAAGATTCGAATAAAGTCTGTAAGAGATTATGCTG 2081
586 LeuAlaMetAlaValHisLysIlePheArgThrPheProGluGlu 602
||||| ||||| ||||| ||||| : : : : : : : : : : : : : : : :
2082 CTGCGCAATATGGAAGTGGCTAAAGATCGCTGAAATTTCCCAAGACA 2131
602 nAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSerAspL 619
||||| ||||| ||||| ||||| : : : : : : : : : : : : : : : :
2132 CGCCCTTCTCTGATACCCGCCCAAGAGAAATGATAAAGGAGC 2181
619 euValGluPheCysAspGlnMetGlyLeuProMetAspValSerSerAla 635
: : : : : : : : : : : : : : : : : : : : : : : : : :
2182 TGCTGAGCAATGTCAGAAATTCGATTTCCATTGACCGAAGACGTCT 2231
636 GlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspLysTyrSe 652
||| ||||| : : : : : : : : : : : : : : : : : : : : : :
2232 GGATTACTTCCACGTCACTTCGCAAGTATCAAGGAAATCCCGTTTGA 2281
652 rLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMetGlnM 669
: : : : : : : : : : : : : : : : : : : : : : : : : :
2282 TATGTGCATTCGACAAGTAATTTCTCATTCACAAATTAAGCCAAATGCAAC 2331
669 etAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluInPheArg 685
: : : : : : : : : : : : : : : : : : : : : : : : : :
2332 AGCGAAATATTTTGCCT...TTCGAGATGCCTTTTCATTTATCAT 2378
686 HisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIleAr 702
||||| ||||| ||||| ||||| : : : : : : : : : : : : : : : :
2379 CATTTTGCCTCAACGTGGATCATATACACATTTTCACATCGCAATTCG 2428
702 gArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGlyT 719
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2429 TCGGTATCCAGATGTGATTGTTCATCGACCAACTTGCTGCTCTGTGT 2478
719 yrSerGluInProAspValGluProAspThrLeuGlnLys..... 732
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2479 ATATGAAGATCTCAAGAGTCTCTCAAGAGATACAGGATTTGTAAGT 2528
733 .....GlnAlaAspHisCys 737
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2529 TTTCAAACACTCAAGAAATTTTTCATTTAAATTTTCAGTGCACCCGCTG 2578
```

```
737 sAsnAspArgMetAlaSerIysArgValGlnGluLeuSerIleGlyL 754
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2579 CAATGCACAAAGCTCGCATCGAGAGCTTCGGATGAGTCTCGCATGC 2628
754 euPhePheAlaValLeuValLysGluSerGlyProLeuGluSerGluAla 770
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2629 TCTATTTTGGAGTTTTCATCCATCAACCGGTCCGATGAAGTCCCAAGCT 2678
771 MetValMetGlyValLeuAsnGlnAlaPheAspValLeuValLeuArgPh 787
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2679 GTTGTCTGTGTAAAGGATCTCATTCGACGTCATCGTTGAGTA 2728
787 edlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArgSerTyrS 804
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2729 TCGAGTTGTGAAGAGATTTATCTGATGT.....GAGTATA 2765
804 erPheGlnLysValGlyLysLys..... 811
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2766 GATTTCAACATGACGCAAAAAAATCCGAATATTTTCAGAAAAATGAAG 2815
812 .....ProGluLeuThrLeuValTrpGluProAs 821
: : : : : : : : : : : : : : : : : : : : : : : : : :
2816 CGAGACTTCAACAAATCGACCGAGAAGCTTACCATTACTGCCACGAGA 2865
821 pAspLeuGluGlu.....GluProThrGlnGlnValIleThrI 834
||||| ||||| : : : : : : : : : : : : : : : : : : : :
2866 TCCGATCCGATCCGGAACCGTGAAGATTTTCGAGCTCAATTCAA 2915
834 lePheSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLys 850
: : : : : : : : : : : : : : : : : : : : : : : : : :
2916 TGTCAACGTGGTTTACGTGATCTC...GTCCGTACAAAAGCATTTGAA 2962
851 TyrSerAlaIleLeuLysArgProGlyLeuGluLys 862
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2963 GTCAGCGGACATTTGTCGCGAGTCTCGAGCAA 2998
seq_name: gb_pl.SPAC2C4
seq_documentation_block:
LOCUS SPAC2C4
DEFINITION S.pombe chromosome I cosmid c2C4.
ACCESSION 299259
VERSION 299259.1 GI:2414611
KEYWORDS 40s ribosomal protein; aminotransferase; cornichon homolog; Homol D
box; lcb2; LTR; mitotic control protein; mRNA processing;
phosphotransferase; proteolipid; ribonuclease II RNB family; rna
binding; rps8; serine palmitoyltransferase 2; serine/threonine
protein kinase; small nuclear ribonucleoprotein; splicing;
tRNA-Met; tRNA-Ser; ubiquitin regulatory domain.
SOURCE
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 35000)
AUTHORS Oliver K. and Harris, D.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 35000)
AUTHORS Bartell, B.G., Rajandream, M.A. and Wood, V.
JOURNAL Submitted (13-SEP-1997) Schizosaccharomyces pombe chromosome I
sequencing project, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S.pombe/)
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
```

number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.  
CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid c2c4 overlapped at the 5' end by cosmid c21E11 and at the 3' end by cosmid c25G10.

#### FEATURES

Source  
1..35000  
/organism="Schizosaccharomyces pombe"  
/strain="972h-"  
/db\_xref="taxon:4896"  
/chromosome="I"  
/map="IR"  
/clone="cosmid c2c4"  
1..214  
/gene="SPAC2c4.01"  
/note="SPAC21E11.07"  
1..214  
/partial  
/gene="SPAC2c4.01"  
/codon\_start=2  
/label="SPAC2c4.01  
/product="hypothetical protein"  
/protein\_id="CAB16361.1"  
/db\_xref="GI:2414612"  
/db\_xref="GI:2414612"  
/translation="PSAPLSIVAKGPEVSRSPGKIIAILKGVGLALVRLQYKLSDL  
ACNGIPFIQNTSIIWLDELSTSPSENS"  
1..1786  
/note="nominal overlap with cosmid SPAC21E11, EM:267999  
S.pombe chromosome 1"  
969..2780  
/gene="lcb2"  
/note="SPAC2c4.02; SPAC21E11.08"  
969..2780  
/gene="SPAC21E11.08"  
/note="SPAC2c4.02, len:603"  
/codon\_start=1  
/label="lcb2"  
/product="serine palmitoyltransferase 2"  
/protein\_id="CAB16362.1"  
/db\_xref="GI:2414613"  
/db\_xref="SWISS-PROT:Q09925"  
/translation="MAQADFVSPTSIDVSEKKEVEFHKKVDHVENPPLSTESAKLEAE  
EVAEKLNSEHLENEFAPITDPTHRRVSKNPDGAELFQDEPSYYVAVYTIYIV  
LIITGHVRDFGKRFHKDDYKLDNDGYAPLYNHFDNFYVRLQHRINDCFSRPTMG  
VPGVIRLMNRYSTDSNSTFKLTGDTSLANVSYNV'LGFAGSHGCPATKVEEAMQRY  
GLTSCSSNAICTGYLGHKEVELTANFVGKPAALVFSQGFSTNATVFTLMCPGSLII  
SDELNTSIFPGARLGSANIRVYKNDMTDLERVILREVISOQPETHRPYSKILVIE  
GLIWMENFCDJPKVVELKKNRYKFLYFIDEAHSIGAIGRGGGICDYGSGISDHWDL  
MGTFISFGAAGGYISATPNIINKLRVTPNPGVYVAESMSPAVLAQKSFLEIMDNP  
TSAGLERLAFNSRYIRLGLKRGIFIENGDDSPVYPILLNFKGINKAFSHEMLKR  
GIAYVVGYPACPLTTSRVRCFSASHNKADMDYFLRACDEVGEKQLQKFSTGAAGED  
VGKTNVEKMKKQGWFKPPRMKIEDVLKHGVHDALQ"  
1668..2552  
/gene="lcb2"  
/note="Match to PF00222 aminotran\_2, Aminotransferases  
class-II Score 305.25"  
2151..2180  
/gene="lcb2"  
/note="PS00599 Aminotransferases class-II  
pyridoxal-phospha te attachment site"  
2655..2678  
/gene="lcb2"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

gene

CDS

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

trna

intron

trna

gene

CDS

gene

CDS





```
|||||
165 TGCAGACCCAGCGCAGTTCGGCACTACGCGCTCAATGTGCCCTGTAC 214
|||||
695 ThrHisPheThrSerProIleArgArgPheAlaAspValIleValHisAr 711
|||||
215 ACACACTTCACCTCGCCATCCCGCGCTTTCGCGAGTCCCTGGTGCACCG 264
|||||
711 gLeuLeuAlaAlaLeuLeuGlyTyrSerGluGlnProAspValGluProA 728
|||||
265 CCTCTGCTGCTGCCGCTTAGGCTATAGGAGCGCACTAGACATGCGCGCCG 314
|||||
728 sPThrLeuGlnLysGlnAlaAspHisCysAsnAspArgMetAlaSer 744
|||||
315 ATACCTTCGAGAAACAGCGGACCACTGTAACGACCGCGCATGGCGTCC 364
|||||
745 LysArgValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuValLy 761
|||||
365 AAGCGCGTGCAGGAGCTCAGTACCACTCTCTCTTGTCTCTGCTGTCAA 414
|||||
761 sGluSerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnG 778
|||||
415 GGAGAGTGGCCCCCTGGAGTCAGACGCCATGGTGAAGGCATCCTGAAGC 464
|||||
778 lAlaPheAspValLeuValLeuArgPheGlyValGlnLysArgIleTyr 794
|||||
465 AAGCCTTCGACGTCTGCTGCTCGCTACGGCGTGCAGAAAGCGCATCTAC 514
|||||
795 CysAsnAlaLeuAlaLeuArgSerTyrSerPheGlnLysValGlyLysLy 811
|||||
515 TGCACAGCACTGCGCCCTGCGGCTCCACCACTTCCAGAGGTTGGCAAGAA 564
|||||
811 sProGluLeuThrLeuValTrpGluProAspAspLeuGluGluProT 828
|||||
565 GCGGAACCTCAGCGTGTCTGGAGCCTCAGGACATGAGCAGGAGCCAG 614
|||||
828 hrcGlnValIleThrIlePheSerLeuValAspValValLeuGlnAla 844
|||||
615 CACACAGCGTCATCACCATCTTCAGCGTGGTGGAGTGGTCTTCGAGGCA 664
|||||
845 GluAlaThrAlaLeuLysTyrSerAlaIleLeuLysArgProGlyLeu.. 860
|||||
665 GAGTCCACAGCCCTCAAGTACAGCGCCATCTCTGAAGCGCGCCAGCCCA 714
|||||
861 .....GluLysAlaSerAspGluGluProG 869
|||||
715 GGGCCACCTGGGCCCTGAGAAGGAGGAGGAGGAGTCTGACGGTGAGCCCG 764
|||||
869 luAsp 870
765 AGGAC 769
```

seq\_name: gb\_pl:AY057479

seq\_documentation\_block:
LOCUS AY057479 2992 bp mRNA linear PLN 07-OCT-2001
DEFINITION Arabidopsis thaliana At2g17510/MJB20.7 mRNA, complete cds.

ACCESSION AY057479
VERSION AY057479.1 GI:15982744
KEYWORDS FLI\_CDNA.
SOURCE thale cress.

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 2992)
Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
Bower,L., Carninci,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu.S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.

TITLE Arabidopsis cDNA clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2992)
AUTHORS Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
Bower,L., Carninci,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu.S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.

TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

COMMENT RIKEN Genomic Sciences Center (RGC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PCEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P., Banh,J., Bower,L.,
Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G.,
Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A.,
Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu.S.,
Davis,R.W., Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.

FEATURES
source
Location/Qualifiers
1..2992
/organism="Arabidopsis thaliana"
/db\_xref="taxon:3702"
/chromosome="2"
/clone="RAFL09-31-H19(RL4592)"
/note="ecotype: Columbia"
1..85
86..2887
/note="putative mitotic control protein dis3"

5'UTR
CDS
/codon\_start=1
/product="At2g17510/MJB20.7"
/protein\_id="AAL09713.1"
/db\_xref="GI:15982745"

/translation="MLQSKVFNKTRGRIGQVREVLRDDIYCGAFSCGSDSSAA
RLSSKIIIVDTNVVLDLLENKADTVVLSVLDVKNRNVNRIILCSNP
ARQYVFSNVHVDYVQAMEKESANDNDRAIRVATLWYOKHLGDTQVLLTNDR
NKRKATEGISAETIEAYVKSIGQPELLDLAQPTNEDITHEDADDSRPSKRLIYQE
HKPSEITAGLRHGIHQGLKVRNFPYEAIVGESIEGIEIIIGSNRNAPDGGDI
HVAVELPRQWODEKALSIAPDEEDDTVHLAPNDVDDAPTNSLSHETSGDKNAAP
VRPSRGVGIIRNWHSYGSLPMSLPAGSGGTALFVSKDRIPKIRINTRQLQAN
LDMRIVAVDSWDRQSPYSGHYVRPICKGKETETEVLIENDYVSPFSQVLA
CLPLPWSVSSDSNVPRQRLHLLVFSVDPGPKDIDALHCTSLPNGNLFGLVHI
ADTVNEVHPGTPLDDEASKRGTSVILVERIDMLPKPLTEDTCSLRADVERLAFSVI
EMSPDAEIIITRTKSIKSSAALSYTEAQRMDSDRLTSLTDLNMTLAKIMRO
RTDRCALTLASAEEKDIDPENNDPLNTGVLLEANQWEEFMLAANVSAGQILK
LPFSCSLRRHPTPREMLEPLRLTAAGLTDVSSKALSDRAVGEDEFFNKL
IRLATRCMTQRTYVFCSDGSLPRLTAAGLTDVSSKALSDRAVGEDEFFNKL
GTYKLPTFQDRPQLTSTVADNLNRRNAAGRASVELYVLIYFRTRPDEARVVK
IRNSGTFVFPKPYGIEGVNLTGKKGAGDWYVDEEKQKTVKMDGSLSYSLQTVKI
HMEVPEPQNPRLQLTL"

misc\_difference 429
/note="compared to genome"
/replaces="g"
2888..2992
3'UTR 837 a 619 c 686 g 850 t
BASE COUNT
ORIGIN





Align seg 1/1 to: AY058669 from: 1 to: 3597

[illegible]

1652 TTCAATAACGGTCTCCGGGATATAAAGCCATTGTGAGCAACGATTTATCGAC 1701

352 CysLeuProGlnSerLeuLeuProTrpThrIleProProAspGluValGlyLy 368

1702 ATTTACTCCCAAGCGGCACCT...CCGATAGACCAAGAGGATCTCGGTCA 1748

368 sArGArGAspLeuArGArGLysAspCysIlePheThrIleAspProSerThrAr 385

1749 ACGAANAAGCTTTAAGAAAGATGTCATATTTACTATAGATCCGATGACTG 1798

385 laArgAspLeuAspAspAlaLeuAlaCysArgArgLeuThrAspGlyThr 401

1799 CTCGCGATTTCGATGACGCGGTTCTCATAGAGAAGTACGGCACAACGAA 1848

402 PheGluValGlyValHisIleAlaAspValSerTyrPheValProGluGlu 418

1849 TATGAGATTGGCTCCATATATCCGATGTTTCGCACCTTCTCATAGAGGA 1898

418 ySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrLeuV 435

1899 CAATGAACGTGATAACTAGTAGAAGACGCTCTAGCTGATTTATTTATG 1948

435 aGlnLysValValProMetLeuProArgLeuLeuCysGluGluLeuCys 451

1949 CCACGAGGTAACTCCACATGCTGCCTCAGTCTCATGCATGCGA...TGT 1995

452 SerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLysLe 468

1996 TCCCTGCTCCCTGGGCACAGCAAGTTGCTTTTCCTGCTCTCTGCGCAT 2045

468 uThrProGluGlyLysIleLeuGluGlu.....TrpPheGlyArgThrI 483

2046 GAACGGGAAGGAGTCATGCTCAAAAGAACCCAGATTTTCCGCCACAG 2095

483 leIleArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIle 499

2096 TCATTAACTCTGCTCGCAATTGGCTTATGAACACGCTCAAAGAATCATC 2145

500 GluAsnProThrGluLysIleProGluGluLeuLeuProIleSerPr 516

2146 GACAACCCCTAATGAAGGTTTACCAGAACGACGTTTCCGACCATCTCGAA 2195

516 oGluHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerI 533

2196 TGGATTAAATCCGATGATATCAGAAATAGGTTGCTATGGCTGCACGACA 2245

533 leAlaLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeu 549

2246 TTGCAAGCTCCATCCGTAAAGACAGCTTTAGACAATAGCGCACTCACAATT 2295

550 AspGlnLeuLysLeuAlaPheThrLeuAspHisGluThrClyLeuProGlu 566

2296 ACAATGCCAAGTACGCTTCTGCTCGATCCCAATTACTGGGAGCGCCGT 2345

566 nGlyCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluP 583

2346 GTCTTTTGAAGTGAGAAAACAGCAGAGCGCTTAACCGCGTGAATGAAGAGT 2395

583 heMetLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhe 599

2396 TTATGCTCTGGCCACACGAGCAGTCGCCGCTTTATACAGCACTCCCTTT 2445

600 ProGluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLe 616

2446 CCGGATATCGTGTATGCGCTAACCAACCCCTCCGCCACTATTAAAGTCCT 2495

616 userAspLeuValGluPheCysAspGlnMetClyLeuProMetAspValS 633

2496 TAAGCTTTAAGGAAAAGCTGCTGCTCTGGGATTCGAATGGCATATTA 2545

633 erSerAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAsp 649

2546 GTCTGCTTAAGCGGCTCCAGAGAGCATGGTCGGTTGTGCATATGACGCG 2595

650 LysTyrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgPr 666  
2596 CCCATCTGCGCATGAATGAATGCTGCTTAAGTCAGCTTTTGATGAAC 2645  
666 oMetGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnGluG 683  
2646 AATGCCCGTGCACATATATTGTTAGCGAGGTAATCGAACCAGCTG 2695

683 InPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSer 699  
2696 ACCTGCGCTACGCTTGTCCATACCATCATACATCATCTTACTTAC 2745

700 ProfileArgPheAlaAspValIleValHisArgLeuLeuAlaAl 716  
2746 CCATCCCGTGGTATCTGATATTTGGTGCACCGACTTTTGGCGCGC 2795

716 aLeuGlyTyrSerGluGlnProAspValGluProAspThrLeuLysG 733  
2796 ACTTAATATTGTTACACCGCTTAACGTTACCCAGATGATTGACAC 2845

733 InAlaAspHisCysAsnAspArgMetAlaSerLysArgValGlnGlu 749  
2846 TAACGAATATTAGCAATGAGCGCACTACATGCGAAAGCGCGGAG 2895

750 LeuSerIleGlyLeuPhePheAlaValLeuValLysGluSerGlyPro 766  
2896 GACTCGTGAATGTTACTTCAACAGCATATGTTCAACACAGCGGCG 2945

766 uGluSerGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspVal 783  
2946 ATACATCGAGCTGTGCTGATTAATTTTTCAGCATATGATGAAC 2995

783 euValLeuArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAla 799  
2996 TTACTCTGGAGTGGCG.....CACGTATAAGC 3024

800 LeuArgSerTyrSerPheGlnLysVal.....GlyLysLysPr 812  
3025 ATT...AACTACAAATGCAAAAGGTCCTCGTAGACACACACGCGT 3071

812 oGluLeuThrLeuValTrpGluProAspAspLeuGluGluProThrC 829  
3072 CAATTACATCTTATTCGCGAG...CGCACTTAACGATCAGCTAGGA 3118

829 InGlnValIleThrIlePheSerLeuValAspValValLeu 842  
3119 AG.....TTGCGAGCTGCTTCGTTGGTCCCATATGCTG 3153

seq\_name: gb\_pl:AC012193

seq\_documentation\_block:

LOCUS AC012193 82454 bp DNA linear PLN 19-JAN-2001  
DEFINITION Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence,  
complete sequence.

ACCESSION AC012193

VERSION AC012193.6 GI:12323286

KEYWORDS HTG.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 82454)  
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,  
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,  
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.

Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence

Unpublished

2 (bases 1 to 82454)

Lin,X. and Kaul,S.

Direct submission

Submitted (21-OCT-1999) The Institute for Genomic Research, 9712

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org  
3 (bases 1 to 82454)  
Town,C.D. and Kaul,S.  
Direct Submission  
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
On Jan 19, 2001 this sequence version replaced gi:12280880.  
Address all correspondence to:at@tigr.org

BAC clone T32E8 is from Arabidopsis thaliana chromosome 1  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.

Genes were identified by a combination of several methods: Gene  
prediction programs including Genscan+ (Chris Burge,  
http://CCR-081.mit.edu/GENSCAN.html), GenemarkHM (Mark Borodovsky,  
http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant  
of GlimmerM, see Mihaela Pertea,  
http://www.tigr.org/softlab/glimmerm\_hum/glimmerm.html, and  
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact  
mpertea@tigr.org), searches of the complete sequence against a  
peptide database and the plant EST database at TIGR  
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to  
indicate the level of evidence for their annotation. Genes with  
similarity to other proteins are named after the database hits.  
Genes without significant peptide similarity but with EST  
similarity are named as unknown proteins. Genes without protein  
or EST similarity, that are predicted by more than two gene  
prediction programs over most of their length are annotated as  
hypothetical proteins. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
Simple repeats are identified by repeatmasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES

source

1..82454  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="T32E8"

misc\_feature

1..189  
/note="overlap with BAC clone T5M16  
[AC010704:103165..103353]."

mRNA

complement(join(<1..69,154..2280,2381..2449,2524..2577,  
2676..3055,3215..53320))

CDS

complement(join(<1..69,154..2280,2381..2449,2524..2577,  
2676..3055,3215..3320))

/gene="T32E8.1"

/codon\_start=1

/product="putative 3'-5' exoribonuclease, 3' partial;  
3320-1"

/protein\_id="AG51632.1"

/db\_xref="GI:12323300"

/translation="MKSSSEQSVRIENGHKRNRPQKNRRSKOSSVPIEDAHVE  
ESLDGRDSRSKAKDSTSSSKQRPNTDEAMRASVAFNSMPMAESGYPRRAS  
PLLSPEVSKQLLSKCPDRACEQPGMNGELFOQIEGSSORIKFSSHWSLDVATEA  
LEKGEAFKALFRVNAHNEYCKIDGVPTDILINGVCSQRAVEGTVVYIKLDPLSL  
WPKMGFVTEAANKPEGTNSPEKDDKAKNGIDVVEGEDGFCSSKSVLGGCAK  
NGVTPSSPDLSDCLGFSCEQKGNCSAVDKLGCILSSFPKRPQTQVAVVKSIVGR  
SIVGLLDVKGWTHYKESDPRKCSPLSDSDEIVOLMPADPRFKLLVPHVLPVSIR  
ARLENIDPLEALVAQAQIVDVEGSPFFVAQITHLFGSGSELEPQNALIYONSVC  
SDFSGSSTLSPRPVPEVPEVORRDKLDLCVLTIDPSTATDLALSVQLPGGF  
FRVGHADVSFVFLPETAIDTAEFRSTSVYLMQRTKISMLPILLSNVGSLSPGADR  
LAFSLINDNREGDVIDRWIGRTIIRSCCKLSYDHAQDIIDGKSDVAENGWALHGSR  
KWCVTYRSVKOLSELSTLRQKFRNGALQENSKPVFLDEHGVYDFVTCRSKGSN  
FLVEFEMLLANTAAEIVISQAYRASLLRHPENPKRLKEFEFGCSKGMGMDLSSS  
GLODSEKIKTNLKKDSVFDILINNYAKPMQLASYFCTGNLKDSVAENGHYALAVP  
LYTHFSLPRFPDVIWVHRAALAAEAELSKQKQTAIDEGRSCTGTGIHFNKDAEES  
IEGKALSVAALKHGVSPSTELSDVAAYCNERNKLAARKVRDADCKLYTFVFLKQKEIF  
PCEARVMNIGSRFMTVYISKL"  
complement(1..3320)  
/gene="T32E8.1"  
/note="similar to ribonuclease II RNB family protein





```

747 aGlnGlu..... 749
|||||
206 TCAGAGATCATGTGTAAGCTTACACTTGGTTGTCTGAAGCAAAAG 157
750 .....LeuSerIleGlyLeuPheAlaVa 758
      ::: ::: :|||:::
156 GAGGTAAAGTAATTCATCACTTCGATCTCTGTGTTAT 107
758 lleuValLysGluSerGly.....ProLeuG 767
      :|||||:|||||:
106 GTTGGTTAGGAACATCAATTATCTTCTCTGCAGATTTTCCGCTG 57
767 luSerGluAlaMetValMetGly 774
      ||:||||:||||:||||
56 AAGCTAGAGTAATGAACTGGGA 34

```

seq\_name: gb\_pr:AF330044

```

seq_documentation_block:
LOCUS      AF330044                7320 bp    mRNA    linear    PRI 02-DEC-2001
DEFINITION Homo sapiens KIAA1008 protein mRNA, complete cds.
ACCESSION  AF330044
VERSION    AF330044.1 GI:17225571
KEYWORDS   human.

```

```

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 7320)
            Royenblum,E., Vahteristo,P., Sandberg,T., Bergthorsson,J.,
            Syrjakoski,K., Weaver,D., Haraldsson,K., Johannsdottir,H.K.,
            Veimanan,P., Nigam,S., Goldberger,N., Robbins,C., Pak,E., Dutra,A.,
            Gillanders,E., Stephan,D., Bailey-Wilson,J., Joo,S., Kaimu,T.,
            Arason,A., Barkardottir,R.B., Nevanlinna,H., Borg,A. and
            Kallioniemi,O.
            A genomic map of a 6 Mb region at 13q21-q22 implicated in cancer
            development: identification and characterization of candidate genes
            unpublished
            2 (bases 1 to 7320)
            Weaver,D., Haraldsson,K., Syrjakoski,K., Bergthorsson,J.,
            Sandberg,T., Vahteristo,P., Rozenblum,E. and Kallioniemi,O.
            Direct Submission
            Submitted (19-DEC-2000) Cancer Genetics Branch, National Institute
            of Human Genome Research, 49 Convent Dr., MSC 4470, Room 4A24,
            Bethesda, MD 20892, USA

```

TITLE Location/Qualifiers

```

1. 7320
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="13"
   /map="13q22"
   /tissue_type="brain; peripheral blood leukocytes"
37. 2913
   /notes="member of ribonuclease II family; similar to
Schizosaccharomyces pombe DIS3, a component of the
exosome"

```

FEATURES

source

```

/codon_start=1
/product="KIAA1008 protein"
/protein_id="AA137479.1"
/db_xref="GI:17225572"
/trans_lation="MLKSKTFLKTRAGGVKMKVIREHYLRDDIGCGAGCAACGAHE
GPALPEQPDPPASSVPOPHYLLPDITVLLHLDIVLEDPAINRVIVLOTVLQEVNRNS
APVTKRIRDTNNOKEHFTTNEHRETVEQEGENANDNDRAIRVAAKWNEHL
KMSADNOLQVITNDNRNKEALEEGIPAFCEYKVLSTANPELIDRLACLSEEG
NEYESKILFSEHLPSKLQGIKSGYLGQTFRASNRYLEATVWIGDSEENKEII
LQGLHLNRVAFHEDIVAVELLPSKVAPSSVVLHDEGNEEDVEKEETERMUKTAV
SEKMLKPTGVWGIIIRNWRPYCGMLSKSDIKESRHELETPADKRIPRIETQAST
LEGRIIVADGWPNRSYPNGHVRNIGLQVKEKETETEVLLLEHDVHPFPFQAVLS
FLPKMWSITEKDMKREDRLHLCITCSVDPGGCTIDDAHLCRELENGLEVGVIAD
VSHFIRPGNALQDOESARRGTTVLCERKIDWVPPELLSKLKDVRDLAFSCIWEM
NHNAEILTKFTKSVINSKASITVAERQIRDSANMNDITTSRLGLNKLAKILKKRR
IEKGAULTSPSEVRFHMDSETHDPIDLQTKRELRETNMSWEEFMILLANISVAKKIHFEF

```

CDS

```

SEHALLRKHPPPPSNYEYLLVKAABSRNLKTDATAESLDOAESPTPPYLNLTLL
RLIATRCMQMAYTFCGMDNDFHHYGLASPIYHTFTSPIRRYADIVHRLLAVALGAD
CTYELTDKHKLADICKNLNFRHMAQASVAFHTLQFTRKSGIVSEAYILFVR
KSNILVPIPKYLEGTFEEDKPNQLIYDDLEISLKIETDVHFHFDKVKVKIMLD
SSNLQHOKIRMSLVEPIGSIPTDTSNMDLNGPKKKMKMLGK*
BASE COUNT      2364 a 1358 c 1521 g 2077 t
ORIGIN

```

```

alignment_scores:
  Quality: 1000.50      Length: 849
  Ratio: 1.962          Gaps: 22
  Percent Similarity: 60.071  Percent Identity: 31.684
alignment_block:
US-09-647-377-9 x AF330044
Align seg 1/1 to: AF330044 from: 1 to: 7320

```

```

47  LysSerIlePheGluThrTyrMetSerLysGluAspValSerGluGlyLe 63
    ||| |||||::: :::: :::: :|||:::
709 AAATAATAATTTTCAGAGCATCTCCCTTAAGTAAGCTACAGCAAGCAT 758
63  uLysArgGlyThrLeuIleGlnGlyValLeuArgIleAsnProLysLysP 80
    :||| ||||| :|||:|||||:|||||: :::: :|||:::
759 AAATCTGGTACATACCTTCAGGAACATTTAGACCTAGCAGGGAATTT 808
80  heHisGluAlaPheIle.....ProSerProAspGlyAspArgAsp 93
    :||| ||||| :||| :||| :|||:
809 ACTTGAAGCTACAGCTATGATGATTCATGCGCAGTGAAGAAATAAGAG 858
94  IlePheIleAspGlyValValAlaAlaArgAsnArgAlaLeuAsnGlyAsp 110
    ||||:||||:||||:||||:||||:||||: :|||:
859 ATATCTTACAGGACTTAACATTTAAACAGAGCTGTTCACGAAGATAT 908
110 uValValLysLeuLeuProGluAspGlnTrpLysAlaValLysProG 127
    ||||:||||:||||:||||:||||:||||: ||| |||:
909 TGTGCTGTGGAGCTTCTCCCAAGAGTCAGTGG.....GTAGCACCAT 952
127 LuSerAsnAspLysGluIleGluAlaThrTyrGluAlaAspIleProGlu 143
    :||| |||||
953 CTCT..... 957
144 GluGlyCysGlyHisHisProLeuGlnGlnSerArgLysGlyTrpSerG 160
957 ..... 957
160 yProAspValIleIleGluAlaGlnPheAspAspSerAspSerGluAspA 177
958 .....GTGGTTTACAT.....GATGAGGTCAAAATGAGAA... 990
177 rgHisGlyAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIle 193
990 ..... 990
194 SerThrProAspArgGlyLysGluAspSerSerThrProValMetLysAs 210
991 .....GATGTGGAGAAAGAGAGACAGACAGAACGATGCTTAAG... 1029
210 pGluAsnThrProIleProGlnAspThrArgGlyLeuSerGluLysSerL 227
1030 .....ACTGCTGTAAGCGGAGAAATGT 1051
227 euGlnLysSerAlaLysValValTyrIleLeuGluLysLysHisSerArg 243
1052 TGAACCTACAGTAGAGTTGTAGGATAATA...AAAGCAATTCGAGA 1098
244 AlaAlaThrGlyTlleLeuLysLeuLeuAlaAspLysAsnSerAspLeuP 260
1099 CCATATTGTGGCATGCTTCCCAAGTCTGCATTAAGAGTCA..... 1140
260 euLysLysTyrAlaLeuPheSerProSerAspHisArgValProArgIle 277
    :||| |||||:||||:||||:||||:||||: |||:

```

```
1141 ....AGAAACATCTCTTTACCTGCTGATAAGAGAAATCCCTCGAATTC 1186
277 yValProLeuLysAspCysProGlnAspPheMetThrArgProLysAsp 293
      : : : : :
1187 GCATAGAAACACGACAG..... 1203
294 PheAlaAsnThrLeu.....PheIleCysArgIleLeuAspTr 306
      : : : : :
1204 ...GCTTCCACATAGAAAGACGAGGAATATTGTTGCTATTGATGTTG 1250
306 pLysGlnAspCysAsnPheAlaLeuGlycInLeuAlaLysSerLeuGlyG 323
      : : : : :
1251 GCCCAGAAATTCAGATATCCAAATGGACACTTGTGAGAAATTTAGGTG 1300
323 InAlaGlyIleLeuProGlnThrGluGlyIleLeuThrGluTyrgly 339
      : : : : :
1301 ATGTTGGAGAGAAAGACGACTGAAACAGAAAGTTTGTACTTGAACACGAT 1350
340 ValAspPheSerAspPheSerSerGluValLeuGluCysLeuProGlnSe 356
      : : : : :
1351 GTTCCCATCAGCCCTTTTCACAGGCTGTCTTAGTTTTCGCCAAG.. 1398
356 rLeuProTrpThrIleProProAspGluValGlyLysArgArgAspLeuA 373
      : : : : :
1399 .ATGCCCTGGAGCATTAAGTAAAGGACATGAAACCGAAGACCTGA 1447
373 iGlyAspCysIlePheThrIleAspProSerThrAlaArgAspLeuAsp 389
      : : : : :
1448 GCATCTGTGATTGTTAGTGTAGACCCACACGAGTGTACTGATATAGAC 1497
390 AspAlaLeuAlaCysArgArgLeuThrAspGlyThrPheGluValGlyVa 406
      : : : : :
1498 GATGCTCTACATTTGTCGAGAACTCGAAATGGAATTTGGAGTTGGTGT 1547
406 IHisIleAlaAspValSerTyrPheValProGluGlySerSerLeuAspL 423
      : : : : :
1548 TCATATTGCTGATGTGAGCCATTTATTAGCCAGGAATGCCTGGATC 1597
423 yValAlaAlaGluArgAlaThrSerValTyrLeuValGlnLysValVal 439
      : : : : :
1598 AAGAAATCAGCAGAGAGGAACTGTTATCTTTGTAAGAGAGGATT 1647
440 ProMetLeuProArgLeuLeuCysGluGluLeuCysSerLeuAsnProMe 456
      : : : : :
1648 GACATGGTTCCAGAGTTGCTAGCTTAACITGTGTTCCITTAATGTA 1697
456 tThrAspLysLeuThrPheSerValIleTrpLysLeuThrProGluGlyL 473
      : : : : :
1698 CTGGACAGGCTGCATTTTCATGTATTGGGAAATGAATCACAATGCTG 1747
473 yIleLeuGluGluTrpPheGlyArgThrIleArgSerCysThrLys 489
      : : : : :
1748 AAATCTTAAACAGAAAGTTTACCAAAAGTGTATTATTAATCAAGGCATCT 1797
490 LeuSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysI 506
      : : : : :
1798 CTGACATATGCTGAAGCTGATGAGAAATGATTACGACAAACATGAAT.. 1845
506 eProGluGluLeuProIleSerProIleSerValIleGluVal 523
      : : : : :
1846 ..... 1852
523 aHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArg 539
      : : : : :
1853 TTACCACTAGTCTCCGTGGACTGAAATAAACTAGCCAAATTTCTGAAGAAA 1902
540 GlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaph 556
      : : : : :
1903 AGAAGGATTGAAAGAGGCTTTGACTCTATCTCTCTGAAAGTTCGATT 1952
556 eThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrGluT 573
      : : : : :
1953 CCACATGGACAGTGAACCTCAGATCCCTATAGATCTGCAGACCAAGGAAC 2002
```

```
573 yrArgAspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsnMet 589
      : : : : :
2003 TTAGGGAACAAATTCATGGTTGAGAAATTTATGTACTTGCCAAATATT 2052
590 AlaValAlaHisLysIlePheArgThrPheProGluGlnAlaLeuLeuAr 606
      : : : : :
2053 TCtGTTGCAAAAAAATTCATGAGAAATTTCTGGAACATGCTCTGCTCG 2102
606 gArgHisProProProGlnThrLysMetLeuSerAspLeuValGluPheC 623
      : : : : :
2103 AAAACATCTCTGCTCCACCTCCATCAATATATGAATTTCTTGTAAAGCAG 2152
623 yAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeuAsn 639
      : : : : :
2153 CCAGGTCAGGAATTTGGAAATTAAGACTGATACAGCCAAGTCTTTGGCT 2202
640 LysSerLeu.....ThrLysThrPheGlyAspAspLysTyrSe 652
      : : : : :
2203 GAGTCTTTGGATCAGGCGCAATCTCTTACTTTTCCA..... 2238
652 rLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMetGln 669
      : : : : :
2239 ....TATCTAAACACTCTGTTGAGAAATATTAGCCACTCCTGCTATGATGC 2284
669 eAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluInPheArg 685
      : : : : :
2285 AAGCTGTGACTCTGTTCTGGAATGGATAATGAT.....TTTCAT 2325
686 HisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIleAr 702
      : : : : :
2326 CACTATGGCTTAGCGTCTCCAATATACACACATTTTACTTCAACCATTAG 2375
702 gArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGly. 718
      : : : : :
2376 AAGATACGAGATGTCATGTTCAICGGCTTTGGCTGGCTATTTGGGG 2425
719 .....TyrSerGluGlnProAspValGluProAspThrLeuGln 731
      : : : : :
2426 CTGACTGTACTTATCCAGAGTTGACAGAC.....AAACAC 2460
732 LysGlnAlaAspHisCys.....AsnAspArgArgMetAlaSerLy 745
      : : : : :
2461 AAGCTTGACAGATATATGTAAAAATCTAAATTTCCGGCACAAGATGGCTCA 2510
745 sArgValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuValLysG 762
      : : : : :
2511 ATATGCCCAACGTCATCAGTGGCTTTTCATACCCAGTTATTTCTCAAAA 2560
762 luSerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnGln 778
      : : : : :
2561 GCAGGAATAGTAGTGAAGAGCCCTATATTTTATTGTAAGAAAGAAAT 2610
779 AlaPheAspValLeuValLeuArgPheGlyValGlnLysArgIleTyrCy 795
      : : : : :
2611 GCCATTGGTATTAAATTCCAAGATGTTGTTAGAGGGACAGTCTTTT.. 2658
795 sasAlaLeuAlaLeuArgSerTyrSerPheGlnLysValGlyLys...L 811
      : : : : :
2659 .....TTGAGAAAGGACAAACCAA 2680
811 ysProGluLeuThrLeuValTrpGluProAspAspLeuGluGluPro 827
      : : : : :
2681 ACCACAGCTTATTTATGATGATGATACCTCCTCCTTAAATAGAAAGAT 2730
828 ThrGlnGlnValIleThrIlePheSerLeuValValLeuGlnAl 844
      : : : : :
2731 ACA.....GTGTTCCATGTATTGTATGAAGTTAAAGTGAATCAATGTT 2774
844 aGluAlaThrAlaLeuLysTyrSerAlaIleLeu..... 855
      : : : : :
2775 AGACATCATTAATCTTCAACATCAGAAAGATCCGAATGTCCCTGCTAGAAC 2824
```

856 ..lvsArgProGlyLeuCluLysAlaSerAspGluCluProGluAsp 870  
::: |||||::: :||| :|||  
2825 CACAGATACAGGAATGAGTCTTCTACTGATCTTCAACATGCG 2871

seq\_name: gb\_in:AY052150

seq\_documentation\_block:  
LOCUS AY052150 3129 bp mRNA linear INV 27-AUG-2001  
DEFINITION Drosophila melanogaster SD10981 full length cDNA.

ACCESSION AY052150

VERSION AY052150.1 GI:15292610

KEYWORDS FLI\_CDNA.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 3129)

Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,

Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,

Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,

Nunoo,J., Pacieb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,

Yu,C., Lewis,S.E., Rubin,G.M. and Celnikier,S.

Direct Submission

Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,

Lawrence Berkeley National Laboratory, One Cyclotron Road,

Berkeley, CA 94720, USA

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory

Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to

sequence clones from Drosophila gene collection 1 (Rubin et al.,

Science 2000). The sequence has been subjected to integrity checks

for sequence accuracy, presence of a polyA tail and contiguity

within 100 kb in the genome. Thus we believe the sequence to

reflect accurately this particular cDNA clone. However, there are

artifacts associated with the generation of cDNA clones that may

have not been detected in our initial analyses such as internal

priming, priming from contaminating genomic DNA, retained introns

due to reverse transcription of unsplined precursor RNAs, and

reverse transcriptase errors that result in single base changes.

For further information about this sequence, including its location

and relationship to other sequences, please visit our Web site

(http://fruitfly.berkeley.edu) or send email to

cdna@fruitfly.berkeley.edu.

Location/Qualifiers

1..3129

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/map="95F7-95F7"

1..3129

/gene="Dis3"

/note="alignment with genomic scaffold AE003747"

/db\_xref="FLYBASE:FBgn0039183"

72..3020

/gene="Dis3"

/note="Longest ORF"

/codon\_start=1

/db\_xref="FLYBASE:FBgn0039183"

/product="SD10981p"

/protein\_id="AAK93574.1"

/db\_xref="GI:15292611"

/translation="MQTREFTRKTNILKIVREHLLDDIGGSELCEKCFONRV

YOLTSOHANKSTLFXPHYLDVLTNNVLDLVEEDLVNRNVLTTLVNLKVRSSA

IYKRFNIVHDTRNRYFVFNHEHKEYADREDETANDRNDRAIRLATKKWDLQIA

SQSKFEKSGRAATRVLLTDDAANKAKAREGILYTTAAEYKSLSEDFELMDKLS

HKTFDSEKGLPQYPLHSMKELLEGLRQNKLLQGTFAQSRNRYLEGTVNVEKEFGI

LIQRESLNRAVDGLIVAVELLPEAWSAPSEIVLEKKNYADEVPSEERAKENML

ESCWVEYDNEANVLKSRPHKSVIKSRAMTYEEAONLIDATQONEIAKSLRNRLA  
KILAKRMDGALVLRASPEIRQVDSETHLEVEYKQREINSWVEEFMLANITVA  
EHIATSEFCAGVLRPRPPPTNFDPVLKVSQGFQVDINSGLSLSHSLDKCKRADN  
PYNTMIRILITTCMMQAVYFISGLQKEEFFHYGLAAPLYHTFTSPRRYSIDIMVHR  
LAAASIGADSTYACQLLDKRSNEELCHNLNRYRHKMAQYARASVALNHLFRGKEEDE  
EGYVLFVRKNALQILPKYKGLTLYLAKSDKDGKGVKSEIVFTNEEDHTQRGC  
DWVHSPDPVTVRLSLDSSNVQHEKILFRLVKPKYIKGVSVEITTEGTSQAVSAPPs  
KKTKKDKKKK"

BASE COUNT 820 a 801 c 841 g 667 t  
ORIGIN

alignment\_scores:

Quality: 998.00 Length: 864

Ratio: 1.912 Gaps: 24

Percent Similarity: 60.417 Percent Identity: 31.019

alignment\_block:

US-09-647-377-9 x AY052150 ..

Align seg 1/1 to: AY052150 from: 1 to: 3129

17 ArgGlyValSerSerValValGlyProSerAlaValGlyAlaSerProGl 33

|||||::: :||| :|||

671 CCGGAGTACCTGAATC.....TCTCGAGGATTTCCACTTC 708

33 YASP..LYSLYSerLysAsnLysSerMetArgGlyLysLysSerLle 49

::: ||| :||| :|||

709 TCGTGATAAGCTCTCGCACAAACATTCGACGCGAGAAGATGCTG 758

50 .....PheGluThrTyrMetSerLysGluAspValSerGluGlyLeu 64

::: :||| :||| :|||

759 CCCAGTACCCGCGCATCTCAGCATGAGGAGCTGCGAAGACTCGG 808

64 sArgGlyThrLeuLeuGlnGlyValLeuArgLleAsnProLysLysPheH 81

::::: ||| :||| :|||

809 CCAAATAAGACTACTACAGGCGACATTTTCAGGCATCCAGGGAATACT 858

81 IsGluAlaPheIleProSerProAspGlyAspArgAspIlePheIleAsp 97

::: :||| :||| :|||

859 TCGAGGCGCGCTCAATGTAGAGAAATTCGAAAGGGGATCCTATCCAA 908

98 GlyValValAlaArgAsnArgAlaLeuAsnGlyAspLeuValVally 114

::: :||| :||| :|||

909 GGACCGAGTCTCTTAATCGGCGCGTGGATGCGACCTGTCGCGCTGGA 958

114 sLeuLeuProGluAspGlnTrpLysAlaValLysProGluSerAsnAspL 131

:||| :||| :|||

959 GCTTTTCCCGAGCGCGAGTGGTCAGCA.....CCCAGTGAAATGTAC 1002

131 ysGluIleGluAlaThrTyrGluAlaAspIleProGluGluGlyCysGly 147

::: :||| :||| :|||

1003 TGGAGGAGAAATGTCTATGCGCATGAAGTGCC..... 1037

148 HisHisProLeuGlnGlnSerArgLysGlyTrpSerGlyProAspValI 164

1037 .....

164 elleGluAlaGlnPheAspAspSerAspSerCluAspArgHisGlyAsnT 181

|||||::: ||| :|||

1038 .....AGCGAGGAAAGGCGCAAGGATG 1059

181 hrSerGlyLeuValAspGlyValLysLysSerIleSerThrProAsp 197

::: :||| :||| :|||

1060 AAAATGAGATCTTGAATCAGTCCGTCGCGCGCTTATCGCA..... 1103

198 ArgGlyLysGluAspSerSerThrProValMetLysAspGluAsnThrPr 214

|||||::: ||| :|||

1104 .....GAGCGCAGCC 1114

214 olleProGlnAspThrArgGlyLeuSerGluLysSerLeuGlnLysSerA 231

|||||::: ||| :|||

1115 C.....ACAG 1119

[illegible]

529 snLeuHISerIleAlaIysGlnLeuArgGlnArgpheValAspGly 545  
1927 ATCTTAATTCGGTTGGCCAAATATCTGAAAGAGCGAGCAATGGATGGA 1976  
546 AlaLeuArgLeuAspGlnLeuLysLeuAlaPheThrLeuAspHisGluThr 562  
1977 GCCTTGTTCCTCCCTCCCGGAGTCGGCTTCAGGTGGACAGCGAGAC 2026  
562 rGlyLeuProGlnGlyCysHisIleTyrGluTyrArgAspSerAsnLysL 579  
2027 ACACGAACCACTCGAGGTGGAGTGAAGACAGATCGCGAGACGAATTCGA 2076  
579 euValGluGluPheMetLeuLeuAlaAsnMetalValAlaHisLysIle 595  
2077 TGGTGGAGGAGTTCATGTTGCTGGCCACATCAGGTGGCCAGCATATC 2126  
596 PheArgThrPheProGluGlnAlaLeuLeuArgHisProProGln 612  
2127 GCACGGAGTTCACGAGTGGCGCTGTTCTGCTGGCCATCTTCGGCCACC 2176  
612 nThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeuP 629  
2177 GCCTACCAATTCGATCGCTGGTAAATCTGCGCGCTACACAGGCTTCC 2226  
629 roMetAspValSerAlaGlyAlaLeuAsnLysSerLeuThrLysThr 645  
2227 AGGTGGACATCAATCGGACCTGGAGCTGTCGCATCTACCTGGACAAGTGT 2276  
646 Phe...GlyAspLysThrSerLeuAlaArgLysGluValLeuThrAs 661  
2277 GTCAAGCGGATATTCCTGACTTAAACAAATGATCCGATCCTCACC... 2324  
661 nMetTyrSerArgProMetGlnMetalLeuTyrPheCysSerGlyMetL 678  
2325 .....ACACGCTGATGTCAGGAGTGTACTTCATCAGCGGCACTC 2367  
678 euGlnAspGlnGluInPheArgHisTyrAlaLeuAsnValProLeuTyr 694  
2368 TGCAG...AAGGAGGAGTTCCTCCACTACGCTGCTGCTGCCATCTAC 2414  
695 ThrHisPheThrSerProIleArgPheAlaAspValIleValHisArg 711  
2415 ACGCACTTTACGTGCGCCATTCGTGTTACTCCGCATATTGTTGATCATCG 2464  
711 gLeuLeuAlaAlaLeuGly.....TyrSerGluGlnProA 724  
2465 CTGCTGGCGCCTCAATCGCGGGGATAGCACTATGCCAGCTATTGG 2514  
724 spValGluProAspThrLeuGlnLysGlnAlaAspHisCysAsnAspArg 740  
2515 AC.....CGAAGTCAATGAGGAGCTGTCTAATCTTT 2549  
741 ArgMetAlaSerLysArgValGln.....GluLe 750  
2550 AACTACCGCCCAAAATGGCTCAGTACGCTGGTCGCGCATCAGTGGCGCT 2599  
750 uSerIleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuG 767  
2600 CAACACGCATCTCTTTCCGC.....GCCAAGGAGGAG.....G 2634  
767 luSerGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeu 783  
2635 ATGAGGAAGGATATTGCTATTGTCGGCAGAAATGCTCTGCAAGTGCTA 2684  
784 ValLeuArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLe 800  
2685 ATCCCCAAGTATGGCTGGAAGGCATTTGTAT.....CT 2719  
800 uArgSerTyrSerPheGlnLysValGlyLys.....L 811  
2720 AAGTCG.....GACAAGGATGCAAGGACGGTGTGGAGCGCGTCA 2760  
811 ysProGluLeuThrLeuValTropIleProAspAspLeuGluGluGluPro 827



```

1154 ..... 1170
294 PheAlaAsnThrLeu.....PheIleCysArgIleLeAspTr 306
1171 ...GCTTCCACATTAGAAGCGGAGAAATTATTGTTCTATTGATGGTTG 1217
306 pLysGluAspCysAsnPheAlaLeuGlyGlnLeuAlaLysSerLeuGlyG 323
1218 GCCCAGAAATCCAGATATCCAAATGGACACTTTGTGAGAAATTTAGGTG 1267
323 lnAlaGlyGluIleGluProGluThrGluGlyIleLeuThrGluIleGly 339
1268 ATGTGTGAGAAAGACAGTGAACAGAAAGTTTGTACTTGAACACAGAT 1317
340 ValAspPheSerAspPheSerGluValLeuGluCysLeuProGlnSe 356
1318 GTTCCCATCAGCCTTTTTCACAGGCTGTTCTTAGTTTCTGCCAAG... 1365
356 rLeuProTrpThrIleProAspGluValGlyLysArgArgAspLeuA 373
1366 .ATGCCCTGGAGCATTTACTGAAAGGACATGAAAAACCGAGAGACCTGA 1414
373 rGlyAspCysIlePheThrIleAspProSerThrAlaArgAspLeuAsp 389
1415 GGCATCTGTGATTGTTAGTGTAGACCACCGAGGATGCTACTGATATAGAC 1464
390 AspAlaLeuAlaCysArgArgLeuThrAspGlyThrPheGluValGlyVa 406
1465 GATGCTCTACATTGTGCGAGAACTCGAAATGGAATTTGGAGTTGGTGT 1514
406 lHisIleAlaAspValSerTyrPheValProGluGlySerSerLeuAspL 423
1515 TCATATTGCTGTGTGAGCCATTTTATTAGCCAGGAAATGCCTGGATC 1564
423 ysValAlaAlaGluArgAlaThrSerValTyrLeuValGlnLysValVal 439
1565 AAGAACTCAGCCAGAGAGAACACTGTGTATCTTTGTGAAAGAGGATT 1614
440 ProMetLeuProArgLeuLeuCysGluGluLeuCysSerLeuAsnProMe 456
1615 GACATGTCTCAGAGTTGTAGCTTAACCTGTGTCTTAAATGTGA 1664
456 tThrAspLysLeuThrPheSerValIleTrpLysLeuThrProGluGlyL 473
1665 CTGGGACAGCTGGGCACTTTTCATGTATTGGGAATGAATCAACATGCTG 1714
473 yslLeuGluGluTrpPheGlyArgThrIleIleArgSerCysThrLys 489
1715 AAATCTTAAAAACGGAAGTTTACCAAAAGTGTATTAAATCAAAAGCATCT 1764
490 LeuSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysIl 506
1765 CTGACATATCTGGAAGCTCAGTTGAGAAATGATTACCAAAATCAATGA 1812
506 eProGluGluGluLeuProProIleSerProGluHisSerValGluV 523
1813 .....GATGATA 1819
523 alHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArg 539
1820 TTACCACTAGTCTCGTGGATCGAATAAATACCAAAATTCAGAGAA 1869
540 GluArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaph 556
1870 AGAAGGATTCAAAAAGGGGCTTTGACTCTATCCTCTCCTGAAGTTCGATT 1919
556 eThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrGluT 573
1920 CCACATGGACAGTGAATCTCAGATCTCTATAGATCTGCAGACCAAGGAAC 1969
573 yrArgaspSerAsnLysLeuValGluGluPheMetLeuLeuAlaAsnMet 589

```

```

1970 TTAGGAAACAAATTCATCGTTGAAGAATTTATTGTTACTTGCCAATATT 2019
590 AlaValAlaHisLysIlePheArgThrPheProGluGlnAlaLeuLeuAr 606
2020 TCTGTTGCAAAAAAATTCATGAGGAATTTCTGAACATGCTCTGCTTCG 2069
606 gArgHisProProGluThrLysMetLeuSerAspLeuValGluPheC 623
2070 AAAACATCCTGCTCCACTCCATCAAAATTTGAAATTTCTTGTTAAGCAG 2119
623 ysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeuAsn 639
2120 CCAGGTCAGGAATTTGAAAATTAAGACTGATACAGCCAGCAAGCTTTGGCT 2169
640 LysSerLeu.....ThrLysThrPheGlyAspAspLysTyrse 652
2170 GAGTCTTTGGATCAGGCCGGAATCTCTACTTTTCCA..... 2205
652 rLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMetGlnM 669
2206 ...TATCTAAACACTCTGTGAGAATATTAGCCACTCGCTGTATGATGC 2251
669 eAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluInPheArg 685
2252 AGCTGTGTACTTCTGTTCTGGAATGGATATGAT.....TTTCAT 2292
686 HisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIleAr 702
2293 CACTATGCTTTAGCGTCTCCAATATACACACTTTTACTTCACCCATTAG 2342
702 gArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGly. 718
2343 AAGATACCCAGATGTCATTGTTCTCGGCTTTTGGCTGTGGCTATTTGGGG 2392
719 .....TyrSerGluGlnProAspValGluProAspThrLeuGln 731
2393 CTGACTGTACTTATCCAGAGTTGACAGAC.....AAACAC 2427
732 LysGlnAlaAspHisCys.....AsnAspArgArgMetAlaSerly 745
2428 AAGCTTGCAATATATGTAATAATCTAAATTTCCGGCACAAATGGCTCA 2477
745 sArgValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuValLysG 762
2478 ATATCCCAACGTCATCAGTGGCTTTTCATACCCAGTTATTCTTCAAAA 2527
762 lUsSerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnGln 778
2528 GCAAAAGGAATAGTAAGTGAAGAGGCTATATTATTGTAAGAAAGAAT 2577
779 AlaPheAspValLeuValLeuArgPheGlyValGlnLysArgIleTyrCy 795
2578 GCCATTGGTATTAAATCCAAAGTATGTTTAGAAGGACAGCTCTTT.. 2625
795 sasnAlaLeuAlaLeuArgSerTyrSerPheGlnLysValGlyLys...L 811
2626 .....TTTCAAGAAAGGACAAACCAA 2647
811 ysProGluLeuThrLeuValTrpGluProAspAspLeuGluGluPro 827
2648 ACCCACAGCTTATTATGATGATGAGATACCCCTCACTTAAATAGAGAAT 2697
828 ThrGlnGlnValIleThrIlePheSerLeuValAspValValLeuGlnAl 844
2698 ACA.....GTGTTCCATGTTATTGATAAGTTAAAGTGAATCATGTT 2741
844 aGluAlaThrAlaLeuLysTyrSerAlaIleLeu..... 855
2742 AGACTCATCTAAATCTTCAACATCAGAAATCCGAATGTCCTCGGTAGAAC 2791
856 ..LysArgProGlyLeuGluLysAlaSerAspGluProGluAsp 870
2792 CACAGATACCAAGGAATAAGCATTTCTACTGTACTTCAACACATGGAC 2838

```



```

95460 CCTGCTCCCTGGCGAAGCAAGTTGGCTTTTTCGGTCTTCTGCGCGATG 95509
469 ThrProGluGlyLysIleLeuGluGlu.....TrpPheGlyArgThrI 483
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95510 AACGGGAAGGAGTCATGCTGCAAAAGAGCGCAGAAATTTTGGCGCAGT 95559
483 eileArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleG 500
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95560 CATTAACCTCTGCTCGCAATTTGCTTATCAACACGCTCAAAAGATCATCG 95609
500 luAsnProThrGluLysIleProGluGluGluLeuProProIleSerPro 516
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95610 ACAACCCCTAATGAACGGTTTACCGAGAGCGACTTTCCGACCATCCTGAAT 95659
517 GluHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerI 533
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95660 GGATTTAATCCGATGATATCAGAAATAGGTTGCTATGGCTGCACGACAT 95709
533 eAlaLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuA 550
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95710 TCGAAGCTCCATCCGTAAGACAGCTTTAGACAATGGCGCACTCAAAATTA 95759
550 spGlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGln 566
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95760 ACAATGGCAAGCTACGCTCTGCTCGATCCAAATTAAGTGGCGAGCCGTTG 95809
567 GlyCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPh 583
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95810 TCTTTTGAAGTGGAGAAACAGCAGAGCGCTAACCGGCTGATTGAAGAGTT 95859
583 eMetLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPheP 600
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95860 TATGCTCTGGCCAAACAGCAGCTGCCCGCTTTTATACACGACTCCTTTC 95909
600 roGluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeu 616
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95910 CGGATATCGCTGCTACGTGCTGTAACCACTCCGCCACTATTAAAGTCCTT 95959
617 SerAspLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSe 633
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
95960 AAGCGTTTAAGGAAAGCTGCTGCTCTGGGATTCGAATGGACTATAG 96009
633 rSerAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspL 650
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96010 CTCGTCTAAGCGCTCAGAGAGCATGGTCGGTTGTGCAATGAGCGGC 96059
650 yTySerSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgPro 666
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96060 CCAATCTCTGCGCAATGAATGCTGCTAAGTCAGCTTTTGATGAAACCA 96109
667 MetGlnMetAla..... 670
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96110 ATGGCCGCTGCAAGCTGAGTCTCAGTTTCAATAATTAATTCCTCCCA 96159
671 .....Leu..TyrPheCysSerGlyMetLeuGlnAspGlnG 682
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
96160 CAATTTCCCTTTAGATATTGTTGACGAGGGTAAATCGGAACCGC 96209
682 uGlnPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrS 699
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96210 TGACTGTGCATACGCTTGTGCTATACCATATACACTCACTTTACTA 96259
699 erProIleArgAspPheAlaAspValIleValHisArg..... 711
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96260 GCCCAATCCGCTGCTGATCTCTGATATTTTGGTGCACCGTAACTGCAC 96309
712 .....Leu.I 713
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96310 ACCTTGAAGTTCGCTCCCGTACTTAAATTTCTCCACTCTAGACTTT 96359
713 euAlaAlaLeuLeuGlyTyrSerGluGlnProAspValGluProAspThr 729
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96360 TGGCGGCGGCACCTTAATATTGTACACCGCTAAACGCTACCCCGAGATG 96409

```

```

730 LeuGlnLysGlnAlaAspHisCysAsnAspArgArgMetAlaSerLysAr 746
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96410 TTGCACACCTTAACGAAATTAGCGAATGAGCGCAAGTACAATGCGAATA 96459
746 gValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuValLysGluS 763
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96460 GGCGGAGAGGACTCTGTAACCTTGTTCAACACGATATGTTCAACACA 96509
763 erGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnGlnAla 779
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96510 AGCAGGCGATATACATGCGAGCTGTGTCATTGAAATATTTTCAGCATATG 96559
780 PheAspValLeuValLeuArgPheGlyValGlnLysArgIleTyrCysAs 796
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96560 ATGAACGTAGTTACTCTGGAGTCGGGC.....CA 96588
796 nAlaLeuAlaLeuArgSerTyrSerPheGlnLysVal.....G 809
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96589 CGTTATAAGCATT...AACTACAAATGCAAAAGGTCCTCTAGACACAC 96635
809 lYlLysLysProGluLeuThrLeuValTrpGluProAspLeuGluGlu 825
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96636 ACGCGCTACCCCAATTACATTCTTATTCGCGAG...CGCAACTTAAGCAG 96682
826 GluProThrGlnGlnValIleThrIlePheSerLeuValAspValValLe 842
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
96683 TCACCTAGGAAG.....TTGCAGCTGCTTCCGTGGTCCCCCATATGCT 96726
842 u 842
96727 G 96727
seq_name: gb_in:AC093121
seq_documentation_block:
LOCUS AC093121
DEFINITION Drosophila melanogaster 3L BAC RP98-11J11 (Roswell Park Cancer
Institute Drosophila BAC Library) complete sequence.
ACCESSION AC093121
VERSION AC093121.1 GI:15148095
KEYWORDS HTG
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 168250)
Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferrier,A., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
Schaefer,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
Direct Submission
Unpublished
2 (bases 1 to 168250)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

```

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Denn, A.L., Ding, X., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgait, D., Edwards, C.C.,  
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Louisgeed, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
 Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,  
 Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,  
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

## TITLE

## JOURNAL

Submitted (11-AUG-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

## COMMENT

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

FEATURES	Location/Qualifiers	alignment_scores:
source	1. 168250 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="3L" /clone="RP98-11J11"	Quality: 972.50 Length: 852 Ratio: 1.918 Gaps: 27 Percent similarity: 59.507 Percent identity: 31.455
BASE COUNT	47670 a 36872 c 37397 g 46311 t	
ORIGIN		
alignment_block:		
US-09-647-377-9 x AC093121/rev ..		
Align seg 1/1 to reverse of: AC093121 from: 1 to: 168250		
66 GlyThrLeuLeuGlnGlyValLeuArgLeuAsnProLysLysPheHisG1 82		
95557 GGGCGCATTTGTAGAGAGGAAATCCGGTTAAACCCGAGAACACAGACA 95508		
82 uAlaPheLeuProSerProAspGlyAspArgAspPheLeuAspGlyV 99		
95507 GGGCTTTATATCATCTCCACTGTCGAGAGGACTCGAACGAGATGGCA 95458		
99 al.....ValAlaArgAsnArgAlaLeuAsnGlyAspLeuVal 111		
95457 TCGTGCTCTTCCAGTGGTAGCGCTATGCTTCGATGGTGACAAAGTA 95408		
112 ValValLysLeuLeu...ProGluAspGlnTrpLysAlaValLysProG1 127		
95407 CGTGCAATTCGTGTATACCCCTGGTGTCTAGGCTAGTTCAAAGACAGCTGA 95358		
127 uSerAsnAspLysLysLeuLeuAlaThrTyrGluAlaAspLeuProGluG 144		
95357 GCGCTCGTCTGGTGAATCTCTGGT..... 95333		
144 LuGlyCysGlyHisHisProLeuGlnInserArg..... 155		
95332 .....GGGAGCCCTCATTTATCTTGGTAAAGTTTATCGCAAACTT 95291		
156 .....LysGlyTrpSerGlyPr 161		
95290 ATAATACTTTCGACTGATGATAATATCTTATAAGCGGATGGTGAGGA.. 95243		
161 oAspValLeuLeuLeuAlaGlnPhe..AspAspSerAspSerGluAspArg 177		
95242 .....GCTTTCGAGCATACCGAGTCCCGAGGCTCT 95212		
178 HisGlyAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSe 194		
95211 GAGTCCGATACCGAACAGCTGGTC..... 95188		
194 rThrProaspArgGlyLysGluaspSerSerThrProValMetLysaspG 211		
95187 .....GTGATATCGTCGG 95175		
211 LuAsnThrProLeuProGlnAspThrArgGlyLeuSerGluLysSerLeu 227		
95174 ACAACTGCCCT..... 95164		
228 GlnLysSerAlaLysValValTyrIleLeuGluLysLysHisSerArgAl 244		
95163 .....AAAGCTTCGTCATAGCAATAAATAACGAACTGAGCTCCGCCA 95120		
244 aAlaThrGlyLeu.....LysLeuLeuAlaAspL 255		
95119 GATCGTGGGACCATATCGTTTACGACCAACCAAACTCTGTGACGACC 95070		
255 ySAsnSerAspLeuPheLysLysTyrAlaLeuPheSerProSerAspHis 271		

95069 AG.....CTATTCTACAAA.....TTCCGACCGTACGATTG 95038  
272 ArgValProArgIleTyrValProLeuLysAspCysProGlnAspPheMe 288  
95037 CGGTGCCCGTGGTCTACGTTCCGAAGGATGCTGTGCGGCCACATCGG 94988  
288 tThrArgProLys...AspPheAlaAsnThrLeuPheIleCysArgIleI 304  
94987 AACAAGCAGCAATAAGACGTGTCGCGCCTCTATACCTTGCCCATATAC 94938  
304 leAspTrpLysLysAspCysAsn...PheAlaLeuGlyGlnLeuAlaLys 319  
94937 TCGAG.....ACAGATTGAAATGGGCACTGCATCGCTGAACCTAATCCAG 94894  
320 SerLeuGlyGlnAlaGlyIleGluProGluThrGluGlyIleLeuTh 336  
94893 CCGGTTGCCCGTGGGTAAATTTGGATGACGAACCTTAAGCAATTCGT 94844  
336 rGluTyrGlyVal...AspPheSerAspPheSerSerGluValLeuGluC 352  
94843 TCATAACGGTCTCGGATATAAGCCATTGAGCAACGATTATCGACA 94794  
352 ysLeuProGlnSerLeuProTrpThrIleProProAspGluValGlyLys 368  
94793 TTTACTCCCGCCGCCACCT...CCGATAAGCCAAAGAGGATCTCGCTCAA 94747  
369 ArgArgAspLeuArgLysAspCysIlePheThrIleAspProSerThrAl 385  
94746 CAAAGACCTTAAGAAGATGTCATATTACTATAGATCGGATGACTGC 94697  
385 aArgAspLeuAspAlaLeuAlaCysArgArgLeuThrAspGlyThrp 402  
94696 TCGGATTGGATGACGCCCTTTCTATAGAAAGCTAGGCGACACGAAT 94647  
402 heGluValGlyValHisIleAlaAspValSerTyrPheValProGluGly 418  
94646 ATGAATGGCTCCATATATCCGATGTTTCGCACCTTCCTCATAGAGGAC 94597  
419 SerSerLeuAspLysValAlaIaGluArgAlaThrSerValTyrLeuVa 435  
94596 AATGAACCTGGATAACATAGTGAAGGAACGCTCTACGCTATTATTAGC 94547  
435 lGlnLysValValProMetLeuProArgLeuLeuCysGluLeuLys 452  
94546 CAACAGGATATCCACATGCTGCTCAGTCCTATGCAATGGA...TGT 94500  
452 erLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLysLeu 468  
94499 CCTGCTCCCTGGGCAAGACAACTTGTCTTTTTCGCTCTTCTGCGGATG 94450  
469 ThrProGluGlyLysIleLeuGluGlu.....TrpPheGlyArgThrI 483  
94449 ACGGGAAGGAGTCATGCTGCAAGAAGAGCCAGAAATTTGCGCGCAGT 94400  
483 eIleArgSerCysThrLysLeuSerTyrAspHisAlaClnSerMetIleG 500  
94399 CATTAACCTCTGTGCGCAATTTGCTTATGAACACGCTCAAAAGATCATCG 94350  
500 luAsnProThrGluLysIleProGluGluLeuProProIleSerPro 516  
94349 ACAACCTTAAGAACGGTTTACGAGACGACTTTCGACCATCCTGAAT 94300  
517 GluHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerI 533  
94299 GGATTTATCCGATATATCAAGAAATAGGGTGTATGGCTGCACGACAT 94250  
533 eAlaLysGlnLeuArgGlnArgPheValAspGlyAlaLeuArgLeuA 550  
94249 TCGAAGCTCCATCGTAAGACAGGTTTAGACAATGGCGCACTCACAATTA 94200  
550 spGlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGln 566  
93223 ACGGCGTACCCCAATTACATCTTATTGCGGAG...CGCAACTTAAGCAG 93277

94199 ACAATGCCAAGCTACGCTTCCTGCTCGATTCOAATTACTGGCGAGCGTTG 94150  
567 GlyCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPh 583  
94149 TCTTTTGAAGTGGAGAAACAGCGAGGCTAACCGCTGATTGAAGAGTT 94100  
583 eMetLeuLeuAlaAsnMetAlaValAlaHisLysIlePheA-gThrPheP 600  
94099 TATGCTCTCGCCACACCGAGCTGCGCGCTTTATACAGCACTCCTTTC 94050  
600 roGluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeu 616  
94049 CGGATATCGCTGCTGCTGCTAACCACCTCGCGCACTTATTAAAGTCCCT 94000  
617 SerAspLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSe 633  
93999 AAGCTTTAGGGGAAAGCTGCTGCTGCTGGGATTCGAATGGACTATAG 93950  
633 rSerAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspL 650  
93949 CTCGCTTAAGCGCTCCAGAGAGCATGGTCCGCTTGTGCAATGAGCGC 93900  
650 ystYrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgPro 666  
93899 CCAATCTGTCGCAATGAATGCCTGAAGTCAAGCTTTGATGAACCA 93850  
667 MetGlnMetAla..... 670  
93849 ATGCGCCGTGCAACGCTGAGTGTCTCAGTTTCAAAATAATTACTTTC 93800  
671 .....Leu...TyrPheCysSerGlyMetLeuGlnAspGlnI 682  
93799 CAATTTCCCGCTTTAGATATTTTGTAGCGAGGTAATAATCGAACC 93750  
682 uGlnPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrs 699  
93749 TGAGCTGTGGCACTACGCTTGTCCATACCATATACACTCATTACTA 93700  
699 erProIleArgArgPheAlaAspValIleValHisArg..... 711  
93659 GCCCAATCGCTCGTATCTCTGATTTTGTGTCACCGCTTAAGTGTCACT 93650  
712 .....Leu.L 713  
93649 ACCTTGAAAGTTGCGCTCCCGTACTTAAATTTCTCCACTCTAGACTTT 93600  
713 euAlaAlaLeuGlyTyrSerGluGlnProAspValGluProAspThr 729  
93599 TGCGCGCGCACTTAAATATTGTACACCGCTAAACGTACCCAGATGAT 93550  
730 LeuGlnLysGlnAlaAspHisCysAsnAspArgArgMetAlaSerLysAr 746  
93549 TTGCACACCTTAACGAAATTAGCGAATGAGCGCAAGTACAATGCGAAAA 93500  
746 gValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuValLysGluS 763  
93499 GGCGGAGAGGACTCTGTAACCTTGTAACCAACGATATGTTCAACA 93450  
763 erGlyProLeuGluSerGluAlaMetValMetGlyValLeuAsnGlnAla 779  
93449 AGCAGGCAATATACATGCGAGCTGTGGTCAATTGAATAATTTACGATATG 93400  
780 PheAspValLeuValLeuArgPheGlyValGlnLysArgIleTyrCysAs 796  
93399 ATGAAGCTAGTTACTCTGAGTCGGGC.....CA 93371  
796 nAlaLeuAlaLeuArgSerTyrSerPheGlnLysVal.....G 809  
93370 CGTTATAAGCAATT...AATACAAATGCAAAAGGTCCTCGTAGACACAC 93324  
809 lyLysLysProGluLeuThrLeuValTrpLeuProAspLeuGluGlu 825  
93323 ACGGCGTACCCCAATTACATCTTATTGCGGAG...CGCAACTTAAGCAG 93277





20301 TGGCGGCGGCACATTAAATATTGTACACGCCCTAAACTACCAGCATGAT 20352  
|||||  
|||

730 LeuGlnLysGlnAlaAspHisCysAsnAspArgArgMetAlaSerLysar 746  
:::  
:::

20251 TTGCACACCTTAACGAATAATTAGCGAATTCAGCGCAAGTACAATCGGAAAA 20202  
|||||  
|||

746 gValcIcnGluLeuSerIlleGlyLeuPhePheAlaValLeuValLysGlus 763  
|||||  
|||

20201 GCCCGGAGAGGAGCTCTGGTAACTTGTTACTTCAAACGATATGTTCCACACA 20152  
|||||  
|||

763 erGlyProLeuGluSerGlnAlaMetValMetGlyValLeuAsnGlnAla 779  
:::  
:::

20151 AGCAGGGCATATACATGCGAGCTGTGTCATTGAAATATTTCAGCATATG 20102  
|||||  
|||

780 PheAspValLeuValLeuArgPheGlyValGlnLysArgIleItyrCysas 796  
|||||  
|||

20101 ATGAACTAGTTACTCTGAGTCCGGC.....CA 20073  
|||||  
|||

796 nAlaLeuAlaLeuArgSertyrSerPheGlnLysVal.....G 809  
|||||  
|||

20072 CGTTATAAGCATT...AACTACAAAATGCAAAAGGTCCTCGTAGACACAC 20026  
|||||  
|||

809 lYlysLysProGluLeuThrLeuValTpGluProAspAspLeuGluGlu 825  
|||||  
|||

20025 ACGCGGTACCCAATTACATTCTTATTCGGAG....CGCAACTTAAGCAG 19979  
|||||  
|||

826 GluProThrGlnGlnValIleThrIlePheSerLeuValAspValVallie 842  
|||||  
|||

19978 TCACCTAGGAAG.....TTGCAGCTGGTTTTCCGTGGTCCCCATATGTCT 19935  
|||||  
|||

842 u 842  
19934 I 19934



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 08:00:23 ; Search time 3767.81 Seconds  
(without alignments)  
17217.499 Million cell updates/sec

Title: us-09-647-377-8  
Perfect score: 3100  
Sequence: 1 cggccgcgcgcgcctcccg.....tgaaaaaaaaaaaaaaaa 3100

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	3100	100.0	3100	6	AX015905	AX015905 Sequence
2	956	30.8	1550	6	AX015899	AX015899 Sequence
3	473	15.3	36901	6	AX015903	AX015903 Sequence
4	473	15.3	38886	6	AX015909	AX015909 Sequence
5	473	15.3	228283	10	AF312994	AF312994 Mus muscu
6	236	7.6	49999	6	AX015902	AX015902 Sequence
7	236	7.6	49999	6	AX015908	AX015908 Sequence
8	116	3.7	49999	6	AX015907	AX015907 Sequence
9	108	3.5	153701	2	AC098189	AC098189 Rattus no
c 10	96	3.1	153701	2	AC098189	AC098189 Rattus no
c 11	57	1.8	63173	2	AC102491	AC102491 Mus muscu
12	49	1.6	1140	6	AX015901	AX015901 Sequence
13	49	1.6	1784	6	AX015910	AX015910 Sequence
14	49	1.6	13330	6	AX015904	AX015904 Sequence
15	49	1.6	49999	6	AX015914	AX015914 Sequence
c 16	49	1.6	175391	2	AC093374	AC093374 Homo sapi
17	49	1.6	181068	9	AC068134	AC068134 Homo sapi
18	47	1.5	139049	9	AC019130	AC019130 Homo sapi
c 19	47	1.5	219775	2	AC105461	AC105461 Homo sapi
20	47	1.5	345102	9	AF312225	AF312225 Homo sapi
21	42	1.4	2189	9	AB066536	AB066536 Macaca fa
22	32	1.0	162842	2	AC024062	AC024062 Homo sapi
23	32	1.0	219775	2	AC105461	AC105461 Homo sapi
24	30	1.0	1739	10	AF126063	AF126063 Mus muscu
25	29	0.9	49999	6	AX015916	AX015916 Sequence
26	29	0.9	158276	2	AC012033	AC012033 Homo sapi
c 27	29	0.9	179816	2	AC074188	AC074188 Homo sapi
28	29	0.9	210301	9	AF307337	AF307337 Homo sapi
29	28	0.9	49999	6	AX015912	AX015912 Sequence
c 30	28	0.9	162842	2	AC024062	AC024062 Homo sapi
31	26	0.8	1592	10	BC001988	BC001988 Mus muscu
32	25	0.8	513	6	AX082887	AX082887 Sequence
c 33	25	0.8	553	6	AX277116	AX277116 Sequence
34	25	0.8	1010	3	AY047550	AY047550 Drosophill
35	25	0.8	1129	9	BC017491	BC017491 Homo sapi
36	25	0.8	1950	8	AF353093	AF353093 Arabidops
37	25	0.8	2617	10	BC023143	BC023143 Mus muscu
38	25	0.8	2732	9	BC014021	BC014021 Homo sapi
39	25	0.8	2852	6	ARL48336	ARL48336 Sequence
40	25	0.8	2904	9	HS090543	U90543 Human butyr
41	25	0.8	162531	9	AC005249	AC005249 Homo sapi
42	25	0.8	275492	2	AC090047	AC090047 Homo sapi
43	24	0.8	565	9	BC022835	BC022835 Homo sapi
44	24	0.8	793	3	RA096081	U96081 Rhipicephal
45	24	0.8	793	6	A93378	A93378 Sequence 17

ALIGNMENTS

RESULT	1	AX015905	3100 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX015905	Sequence 8 from Patent WO9950284.				
DEFINITION	AX015905					
ACCESSION	AX015905					
VERSION	AX015905.1	GI:10041648				
KEYWORDS						
SOURCE		house mouse.				
ORGANISM		Mus musculus				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
TITLE		1 (bases 1 to 3100)				
JOURNAL		Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.				
FEATURES		Nucleic acid molecules which code proteins influencing bone development				
source		1. .3100				
CDS		/organism="Mus musculus"				
		/db_xref="taxon:10090"				
		125. .2737				







QY 2612 caggtcatcaccatctcagcctgtgtgatgtgtcctcagcagagcagagccacagccctc 2671  
|||||  
Db 20064 CAGGTCAATACCAATCTTACGCTGGTGATGTGGTCTCTGAGCGAGAGCCACAGCCCTC 20123  
|||||  
QY 2672 aagtacagtgctatcctgaagcaccagggcctggagaagcgctctgatgagagcctgag 2731  
|||||  
Db 20124 AAGTACAGTGCTATCTGAAGCGACAGGCGCTGAGAAAGCGCTCTGATGAGAGCCCTGAG 20183  
|||||  
QY 2732 gactgaatcagcaccagcagcctgtcctgcccctaccctgctggcttttaggaata 2791  
|||||  
Db 20184 GACTGAATGCTAGCCACAGCCAGCCCTGCTGCCCTACCCCTGCTGGCTTTTAGGAATA 20243  
|||||  
QY 2792 ggaacctttgacaccaaaggggatttttaatttgggttttaacaaactcagggggtttgtt 2851  
|||||  
Db 20244 GGACCTTTTGACACCAAAAGGGGATTTTAAATTTGGTTTAAACACTCAGGGGTTGTGT 20303  
|||||  
QY 2852 ttattttatttctctttattttacttttgcagctcagcttcttaaatgaactggaag 2911  
|||||  
Db 20304 TTATTTTATTTTCTCTTTTATTTTACTTTTGCAGCTCAGTTTAAATGAACCTGGAAG 20363  
|||||  
QY 2912 ttagggtcagggcaggggatgctgagggcctgctgtcctcctcagcagagaggtatc 2971  
|||||  
Db 20364 TTAGGGTCTAGGGCAGGGGATGCTGAGGCCCTGGCCCTGTGCTTCCCTGAGCAGAGGATC 20423  
|||||  
QY 2972 ccagtcctcctgagcagcagcccgcttaccagagcagccactgcccctccctgcc 3031  
|||||  
Db 20424 CCAGTCTCTCTGGCAGGCAGCCCGCTTCTACAGGCGACCTGCTTCCCTGCTGCC 20483  
|||||  
QY 3032 aggaatcgggggtttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 3084  
|||||  
Db 20484 AGGAATGGGGGTTTTCAGCAATCAGTGTCTATGGAATAAAATCAAGTGTGAA 20536  
|||||  
RESULT 5  
AF312994 228283 bp DNA linear ROD 17-JAN-2002  
LOCUS AF312994 Mus musculus chromosome 1 clone MML, complete sequence.  
DEFINITION AF312994  
ACCESSION AF312994  
VERSION AF312994.1 GI:18182288  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 228283)  
Rump,A., Hess,J., Galgoczy,P., Wirth,T. and Rosenthal,A.  
TITLE Genomic sequence from mouse chr. 1  
JOURNAL Unpublished  
AUTHORS Rump,A., Hess,J., Galgoczy,P., Wirth,T. and Rosenthal,A.  
REFERENCE 2 (bases 1 to 228283)  
Submitted (12-OCT-2000) Genome Analysis, Institute of Molecular  
TITLE Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
JOURNAL  
FEATURES  
source  
1..228283  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/clone="MML"  
BASE COUNT 63688 a 49622 c 49154 g 65819 t  
ORIGIN  
Query Match 15.3%; Score 473; DB 10; Length 228283;  
Best Local Similarity 100.0%; Pred. No. 1.3e-252;  
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2612 caggtcatcaccatctcagcctgtgtgatgtgtcctcagcagagcagagccacagccctc 2671  
|||||  
Db 209461 CAGGTCAATACCAATCTTACGCTGGTGATGTGGTCTCTGAGCGAGAGCCACAGCCCTC 209520  
|||||  
QY 2672 aagtacagtgctatcctgaagcaccagcctggagaagcgctctgatgagagcctgag 2731  
|||||

Db 209521 AAGTACAGTGCTATCTTAAAGCAGCAGCCCTGAGAGGCGTCTGATGAGAGCCTGAG 209580  
|||||  
QY 2732 gactgaatcagcaccagcagcctgtcctgcccctaccctgctggcttttaggaata 2791  
|||||  
Db 209581 GACTGAATGCTAGCCACAGCCAGCCCTGCTGCCCTACCCCTGCTGGCTTTTAGGAATA 209640  
|||||  
QY 2792 ggaacctttgacaccaaaggggatttttaatttgggttttaacaaactcagggggtttgtt 2851  
|||||  
Db 209641 GGACCTTTTGACACCAAAAGGGGATTTTAAATTTGGTTTAAACACTCAGGGGTTGTGT 209700  
|||||  
QY 2852 ttattttatttctctttattttacttttgcagctcagcttcttaaatgaactggaag 2911  
|||||  
Db 209701 TTATTTTATTTTCTCTTTTATTTTACTTTTGCAGCTCAGTTTAAATGAACCTGGAAG 209760  
|||||  
QY 2912 ttagggtcagggcaggggatgctgagggcctgctgtcctcctcagcagagaggtatc 2971  
|||||  
Db 209761 TTAGGGTCTAGGGCAGGGGATGCTGAGGCCCTGGCCCTGTGCTTCCCTGAGCAGAGGATC 209820  
|||||  
QY 2972 ccagtcctcctgagcagcagcccgcttctaccagggcagccactgcccctccctgcc 3031  
|||||  
Db 209821 CCAGTCTCTCTGGCAGGCAGCCCGCTTCTACAGGCGACCCACTGCTTCCCTGCC 209880  
|||||  
QY 3032 aggaatcgggggtttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 3084  
|||||  
Db 209881 AGGAATGGGGGTTTTCAGCAATCAGTGTCTATGGAATAAAATCAAGTGTGAA 209933  
|||||  
RESULT 6  
AX015902 49999 bp DNA linear PAT 07-SEP-2000  
LOCUS AX015902 Sequence 5 from Patent WO9950284.  
DEFINITION AX015902  
ACCESSION AX015902  
VERSION AX015902.1 GI:10041645  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 49999)  
Aligner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.  
TITLE Nucleic acid molecules which code proteins influencing bone  
JOURNAL development  
AUTHORS Patent: WO 9950284-A 5 07-OCT-1999;  
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP  
ANDREAS (DE); WIRTH THOMAS (DE)  
FEATURES  
source  
1..49999  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
BASE COUNT 13210 a 11814 c 10825 g 14150 t  
ORIGIN  
Query Match 7.6%; Score 236; DB 6; Length 49999;  
Best Local Similarity 100.0%; Pred. No. 7.2e-120;  
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1542 agatccttgaagtggttggccgcactatcatccgttcttgcacaaactgagtaag 1601  
|||||  
Db 8518 AGATCCTTGAAGAGTGGTTTGGCGGCACATCATCGTCTTCTTGACCAAACTGAGCTAGC 8577  
|||||  
QY 1602 accatgccagagcatgatcgaataaaactcgaagatccctcaggaagagctccccc 1661  
|||||  
Db 8578 ACCATGCCAGAGCATGATCGAAATCCAACTGAGAGATCCCTGAGGAAGAGTTTCCCC 8637  
|||||  
QY 1662 caattcttcagagcacagcgtcagaggtgcaaccaggcagctcctgaacctgacagca 1721  
|||||  
Db 8638 CAATTTCTCAGAGCACAGCGTCGAGGAGGTGCACACGAGCAGTCTCTGAACCTGCACAGCA 8697  
|||||  
QY 1722 ttgcaagaactccgcagcagccttttagatggcgactccgttttagatcag 1777  
|||||  
Db 8698 TTGCAAGCAACTCCGCCGCCGCGCTTTGTAGATGGCGCACTCCGTTTAGATCAG 8753  
|||||

```

RESULT 7
AX015908
LOCUS AX015908 4999 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 11 from Patent WO9950284.
ACCESSION AX015908
VERSION AX015908.1 GI:10041651
KEYWORDS
SOURCE
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 4999)
AUTHORS Aigner, T., Hess, J., Rosenthal, A., Rump, A. and Wirth, T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 11 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Location/Qualifiers
source 1..4999
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 13135 a 11787 c 10868 g 14209 t
ORIGIN

Query Match 7.6%; Score 236; DB 6; Length 4999;
Best Local Similarity 100.0%; Pred. No. 7.2e-120;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1542 agatccttgagagtggttggccgcactatcccgcttctgaccacactgagctacg 1601
|||||
Db 10503 AGATCCTTGAAGAGTGGTTGGCCGCCTACTATCCGTCCTTGCCACCAACTGAGCTACG 10562

QY 1602 accatgccgagcagatgatcgaaaaatccaaactgagaagatccctgaggaagagttcccc 1661
|||||
Db 10563 ACCATGCCGAGAGCATGTCGAAAATCCAACTGAGAAGATGCCCTGAGGAAGAGCTTCCCC 10622

QY 1662 caatttctccagagcagcgcgtcgaggaggtgcaccagggcagtcctgaaacctgcacagca 1721
|||||
Db 10623 CAAATTCTCCAGAGCAGCAGCGTCGAGGAGGTGCACCAGGCAGTCTGAACTGCACAGCA 10682

QY 1722 ttgcaagcaactccgccccagccttctgtagatgagcgcactccgctttagatcag 1777
|||||
Db 10683 TTGCANAGCAACTCCGCCGCCAGCGCTTTGTAGATGGCGCACTCCGCTTTAGATCAG 10738

RESULT 8
AX015907
LOCUS AX015907 4999 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 10 from Patent WO9950284.
ACCESSION AX015907
VERSION AX015907.1 GI:10041650
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 4999)
AUTHORS Aigner, T., Hess, J., Rosenthal, A., Rump, A. and Wirth, T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 10 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Location/Qualifiers
source 1..4999
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 14550 a 10214 c 10719 g 14516 t
ORIGIN

```

```

Query Match 3.7%; Score 116; DB 6; Length 4999;
Best Local Similarity 100.0%; Pred. No. 1.1e-52;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1322 ggcaccttgaaagtggcgctccacatgcgcatgtagtgagttacttcttgcaggatcc 1381
|||||
Db 30110 GGCACCTTGAAGTGGCGCTCCACATCGCGCATGTGAGTACTTGTCTCTGAGGATCC 30169

QY 1382 tctttgataaagtgcgtgctgagagccacagtcgtctacttgggtccagaaggt 1437
|||||
Db 30170 TCTTTGATAAAGTAGCTGCTGAGAGAGCCACAAGTGTCTACTTGGTCCAGAAGGT 30225

RESULT 9
AC098189
LOCUS AC098189 153701 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-145D14, *** SEQUENCING IN PROGRESS
***, 61 unordered pieces.
ACCESSION AC098189
VERSION AC098189.3 GI:17973688
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 153701)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barbaria, J.,
Benton, J., Blmage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M., Davis, C.,
Davy-Carroll, D., Dederich, P.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drapet, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Hartis, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S.,
Joudan, S., Karlsson, E., Kelly, S., Khan, B., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogun, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Roife, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,
Sisson, H., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watling, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 153701)
AUTHORS Worley, K.C.
Direct Submission
TITLE
JOURNAL Submitted (23-OCT-2001) Human Genome Sequencing Center, Department

```

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:17064556.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GIHF  
Center clone name: CH230-145D14  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList  
Consensus quality: 123669 bases at least Q40  
Consensus quality: 129396 bases at least Q30  
Consensus quality: 135066 bases at least Q20  
Estimated insert size: 124661: sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 61 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1 8487: contig of 8487 bp in length  
\* 8488 8587: gap of unknown length  
\* 8588 12859: contig of 4272 bp in length  
\* 12860 12959: gap of unknown length  
\* 12960 17729: contig of 4770 bp in length  
\* 17730 17829: gap of unknown length  
\* 17830 23516: contig of 5687 bp in length  
\* 23517 23616: gap of unknown length  
\* 23617 27189: contig of 3573 bp in length  
\* 27190 27289: gap of unknown length  
\* 27290 31122: contig of 3833 bp in length  
\* 31123 31222: gap of unknown length  
\* 31223 35679: contig of 4457 bp in length  
\* 35680 35779: gap of unknown length  
\* 35780 39069: contig of 3290 bp in length  
\* 39070 39169: gap of unknown length  
\* 39170 43423: contig of 4254 bp in length  
\* 43424 43523: gap of unknown length  
\* 43524 47921: contig of 4398 bp in length  
\* 47922 48021: gap of unknown length  
\* 48022 51363: contig of 3342 bp in length  
\* 51364 51463: gap of unknown length  
\* 51464 54539: contig of 3076 bp in length  
\* 54540 54639: gap of unknown length  
\* 54640 57364: contig of 2725 bp in length  
\* 57365 57464: gap of unknown length  
\* 57465 61911: contig of 4447 bp in length  
\* 61912 62011: gap of unknown length  
\* 62012 66486: contig of 4475 bp in length  
\* 66487 66586: gap of unknown length  
\* 66587 69977: contig of 3391 bp in length  
\* 69978 70077: gap of unknown length  
\* 70078 73099: contig of 3022 bp in length  
\* 73100 73199: gap of unknown length  
\* 73200 76223: contig of 3024 bp in length  
\* 76224 76323: gap of unknown length  
\* 76324 78464: contig of 2041 bp in length  
\* 78465 78464: gap of unknown length  
\* 78465 80632: contig of 2168 bp in length  
\* 80633 80732: gap of unknown length  
\* 80733 83076: contig of 2344 bp in length  
\* 83077 83176: gap of unknown length  
-----  
86186: contig of 3010 bp in length  
86187 gap of unknown length  
89099: contig of 2813 bp in length  
89100 gap of unknown length  
91514: contig of 2315 bp in length  
91515 gap of unknown length  
93490: contig of 1876 bp in length  
93491 gap of unknown length  
95001: contig of 1411 bp in length  
95101: gap of unknown length  
97069: contig of 1968 bp in length  
97169: gap of unknown length  
97170 contig of 2053 bp in length  
99222: gap of unknown length  
102096: contig of 2774 bp in length  
102196: gap of unknown length  
103303: contig of 1107 bp in length  
103403: gap of unknown length  
105586: contig of 2183 bp in length  
105686: gap of unknown length  
107855: contig of 2169 bp in length  
107955: gap of unknown length  
110106: contig of 2151 bp in length  
110206: gap of unknown length  
112025: contig of 1819 bp in length  
112125: gap of unknown length  
113623: contig of 1498 bp in length  
113723: gap of unknown length  
115968: contig of 2245 bp in length  
116088: gap of unknown length  
118090: contig of 2022 bp in length  
118190: gap of unknown length  
118191 119318: contig of 1128 bp in length  
119319 119418: gap of unknown length  
120531: contig of 1113 bp in length  
120631: gap of unknown length  
122162: contig of 1531 bp in length  
122262: gap of unknown length  
124304: contig of 2042 bp in length  
124404: gap of unknown length  
126716: contig of 2312 bp in length  
126816: gap of unknown length  
127986: contig of 1170 bp in length  
128086: gap of unknown length  
129847: contig of 1761 bp in length  
129848 129947: gap of unknown length  
131590: contig of 1643 bp in length  
131690: gap of unknown length  
131691 133025: contig of 1335 bp in length  
133125: gap of unknown length  
134625: contig of 1500 bp in length  
134725: gap of unknown length  
135806: contig of 1081 bp in length  
135906: gap of unknown length  
137024: contig of 1118 bp in length  
137124: gap of unknown length  
138379: contig of 1255 bp in length  
138479: gap of unknown length  
139623: contig of 1144 bp in length  
139723: gap of unknown length  
140842: contig of 1119 bp in length  
140942: gap of unknown length  
142718: contig of 1776 bp in length  
142818: gap of unknown length  
143800: contig of 1162 bp in length  
144080: gap of unknown length  
145323: contig of 1243 bp in length  
145423: gap of unknown length  
146688: contig of 1265 bp in length  
146788: gap of unknown length  
-----  
Query Match 3.5%: Score 108; DB 2; Length 153701;  
Best Local Similarity 100.0%; Pred. No. 3.4e-48;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2170 gcagttccggcattatgctctcaacgttcctccctctacacacacttcacctctccatccg 2229  
|||||

Db 132036 GCAGTTCGGCATTATGCTCTCAACGTTCCCTCTACACACTTCACCTCCCATCCG 132095  
|||||

Qy 2230 ccgctttgctgagctcatagtcacccctccctgctgctgtctgtggg 2277  
|||||

Db 132096 CCCTTGTGCTGAGCTCATAGTCACCGCTCTGCTGCTGCTG 132143  
|||||

# RESULT 10

AC098189/c

LOCUS

DEFINITION

AC098189

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC098189 153701 bp DNA linear HTG 20-DEC-2001

Rattus norvegicus clone CH230-145D14, \*\*\* SEQUENCING IN PROGRESS

\*\*\* 61 unordered pieces.

AC098189.3 GI:17973688

HTG; HTGS\_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 153701)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carlton,T.F.,

Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Covle,M.D., Dathore,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.K., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Hochs,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,

Hollins,B., Honsi,F., Howard,S., Huber,J., Huiy,J., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Katsic,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louisseg,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mayhew,E., McLeod,M.P., Meador,M.,

Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,

Ogduh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,

Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,

Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Tamerisa,K.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Watlington,S., Williams,G., Williamson,A., Wleczek,R., Woodson,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

## TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission

Unpublished

2 (bases 1 to 153701)

Worley,K.C.

Submitted (23-OCT-2001)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17064556.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GRHF  
Center clone name: CH230-145D14  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList  
Consensus quality: 123669 bases at least Q40  
Consensus quality: 129396 bases at least Q30  
Consensus quality: 135066 bases at least Q20  
Estimated insert size: 124661; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 61 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 8487: contig of 8487 bp in length  
8587: gap of unknown length  
8588: contig of 4272 bp in length  
12860: gap of unknown length  
12860: contig of 4770 bp in length  
17730: gap of unknown length  
17829: contig of 5687 bp in length  
23516: gap of unknown length  
23517: contig of 3573 bp in length  
23617: gap of unknown length  
27189: contig of 3573 bp in length  
27190: gap of unknown length  
27290: contig of 3833 bp in length  
31122: gap of unknown length  
31223: contig of 4457 bp in length  
35779: gap of unknown length  
35780: contig of 3290 bp in length  
39069: gap of unknown length  
39070: contig of 4254 bp in length  
39170: gap of unknown length  
43423: contig of 4254 bp in length  
43523: gap of unknown length  
43524: contig of 4398 bp in length  
47921: gap of unknown length  
48021: contig of 3342 bp in length  
48022: gap of unknown length  
51363: contig of 3076 bp in length  
51364: gap of unknown length  
54539: contig of 2725 bp in length  
54540: gap of unknown length  
57364: contig of 4447 bp in length  
57365: gap of unknown length  
57465: contig of 4447 bp in length  
61911: gap of unknown length  
62012: contig of 4475 bp in length  
65487: gap of unknown length  
65587: contig of 3391 bp in length  
69978: gap of unknown length  
70077: contig of 3022 bp in length  
73100: gap of unknown length  
73199: contig of 3024 bp in length  
76223: gap of unknown length  
76324: contig of 2041 bp in length  
78364: gap of unknown length  
78365: contig of 2168 bp in length  
80632: gap of unknown length  
80633: contig of 2344 bp in length  
80733: gap of unknown length  
83077: contig of 3010 bp in length  
83177: gap of unknown length  
86187: contig of 2813 bp in length  
89099: gap of unknown length  
89099: contig of 2813 bp in length

\* 89100 89199: gap of unknown length  
\* 89200 91514: contig of 2315 bp in length  
\* 91515 93614: gap of unknown length  
\* 93615 93490: contig of 1876 bp in length  
\* 93491 93590: gap of unknown length  
\* 93591 95001: contig of 1411 bp in length  
\* 95002 95101: gap of unknown length  
\* 95102 97069: contig of 1968 bp in length  
\* 97070 97169: gap of unknown length  
\* 97170 99222: contig of 2053 bp in length  
\* 99223 99222: gap of unknown length  
\* 99223 102096: contig of 2774 bp in length  
\* 99223 102096: gap of unknown length  
\* 102097 102196: gap of unknown length  
\* 102197 103303: contig of 1107 bp in length  
\* 103304 103403: gap of unknown length  
\* 103404 105586: contig of 2183 bp in length  
\* 105587 105686: gap of unknown length  
\* 105687 107855: contig of 2169 bp in length  
\* 107856 107955: gap of unknown length  
\* 107956 110106: contig of 2151 bp in length  
\* 110107 110206: gap of unknown length  
\* 110207 112025: contig of 1819 bp in length  
\* 112026 112125: gap of unknown length  
\* 112126 113623: contig of 1498 bp in length  
\* 113624 113723: gap of unknown length  
\* 113724 115968: contig of 2245 bp in length  
\* 115969 116068: gap of unknown length  
\* 116069 118090: contig of 2022 bp in length  
\* 118091 118190: gap of unknown length  
\* 118191 119318: contig of 1128 bp in length  
\* 119319 119418: gap of unknown length  
\* 119419 120531: contig of 1113 bp in length  
\* 120532 120631: gap of unknown length  
\* 120632 122616: contig of 1531 bp in length  
\* 122616 122626: gap of unknown length  
\* 122626 124304: contig of 2042 bp in length  
\* 124305 124404: gap of unknown length  
\* 124405 126716: contig of 2312 bp in length  
\* 126717 126816: gap of unknown length  
\* 126817 127986: contig of 1170 bp in length  
\* 127987 128086: gap of unknown length  
\* 128087 129847: contig of 1761 bp in length  
\* 129848 129947: gap of unknown length  
\* 129948 131590: contig of 1643 bp in length  
\* 131591 131690: gap of unknown length  
\* 131691 133025: contig of 1335 bp in length  
\* 133026 133125: gap of unknown length  
\* 133126 134625: contig of 1500 bp in length  
\* 134626 134725: gap of unknown length  
\* 134726 135806: contig of 1081 bp in length  
\* 135807 135906: gap of unknown length  
\* 135907 137024: contig of 1118 bp in length  
\* 137025 137124: gap of unknown length  
\* 137125 138379: contig of 1255 bp in length  
\* 138380 138479: gap of unknown length  
\* 138480 139623: contig of 1144 bp in length  
\* 139624 139723: gap of unknown length  
\* 139724 140842: contig of 1119 bp in length  
\* 140843 140942: gap of unknown length  
\* 140943 142718: contig of 1776 bp in length  
\* 142719 142818: gap of unknown length  
\* 142819 143980: contig of 1162 bp in length  
\* 143981 144080: gap of unknown length  
\* 144081 145323: contig of 1243 bp in length  
\* 145324 145423: gap of unknown length  
\* 145424 146688: contig of 1265 bp in length  
\* 146689 146788: gap of unknown length

Query Match 3.1%; Score 96; DB 2; Length 153701;  
Best Local Similarity 99.3%; Pred. No. 1.8e-41;  
Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1895 gTggcccaagaatcttcgcagacattccctgagcaggccctgctgcgcggcatcccca 1954

Db 77838 GTGCCCCAAGATCTCCGCACCTTCCTCCGTGACGAGGCCCTGCTGCGCAGCATCCCCCA 77779  
QY 1955 ccacagacgaagatctcagtgaccctggtggagttcttgaccagatggagctgcccatg 2014  
Db 77778 CCACAGCAGGAATGCTCAGTGACCTGGTGGAGTCTTGACACGATGGGCTGCCCATG 77719  
QY 2015 gatgcagctctgcaggggccctaaat 2041  
Db 77718 GATGTCAGCTCTGCAGGGGCCCTAAAT 77692  
  
RESULT 11  
AC102491/c  
LOCUS Mus musculus clone RP24-529L13, LOW-PASS SEQUENCE SAMPLING. 63173 bp DNA linear HTG 23-NOV-2001  
DEFINITION AC102491  
ACCESSION AC102491  
VERSION AC102491.1 GI:17061577  
KEYWORDS HTG: HTGS\_PHASE0.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 63173)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Mus musculus, clone RP24-529L13  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 63173)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,  
Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Rhmann,S., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L18907  
Center clone name: 529\_L13  
-----

\* NOTE: This record contains 81 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 746 845: gap of 100 bp  
\* 1486: contig of 641 bp in length  
\* 1586: gap of 100 bp  
\* 2340: contig of 754 bp in length  
\* 2341 2440: gap of 100 bp  
\* 2441 3115: contig of 675 bp in length  
\* 3116 3215: gap of 100 bp  
\* 3216 3906: contig of 691 bp in length  
\* 3907 4006: gap of 100 bp  
\* 4007 4706: contig of 700 bp in length  
\* 4707 4806: gap of 100 bp  
\* 4807 5488: contig of 682 bp in length  
\* 5489 5588: gap of 100 bp  
\* 5589 6265: contig of 677 bp in length  
\* 6266 6365: gap of 100 bp  
\* 6366 7025: contig of 660 bp in length  
\* 7026 7125: gap of 100 bp  
\* 7126 7787: contig of 662 bp in length  
\* 7788 7887: gap of 100 bp  
\* 7888 8527: contig of 640 bp in length  
\* 8528 8627: gap of 100 bp  
\* 8628 9317: contig of 690 bp in length  
\* 9318 9417: gap of 100 bp  
\* 9418 10070: contig of 653 bp in length  
\* 10071 10170: gap of 100 bp  
\* 10171 10836: contig of 666 bp in length  
\* 10837 10936: gap of 100 bp  
\* 10937 11634: contig of 698 bp in length  
\* 11635 11734: gap of 100 bp  
\* 11735 12425: contig of 691 bp in length  
\* 12426 12525: gap of 100 bp  
\* 12526 13220: contig of 695 bp in length  
\* 13221 13320: gap of 100 bp  
\* 13321 14017: contig of 697 bp in length  
\* 14018 14117: gap of 100 bp  
\* 14118 14808: contig of 691 bp in length  
\* 14809 14908: gap of 100 bp  
\* 14909 15565: contig of 657 bp in length  
\* 15566 15665: gap of 100 bp  
\* 15666 16353: contig of 688 bp in length  
\* 16354 16453: gap of 100 bp  
\* 16454 17123: contig of 670 bp in length  
\* 17124 17223: gap of 100 bp  
\* 17224 17911: contig of 688 bp in length  
\* 17912 18011: gap of 100 bp  
\* 18012 18692: contig of 681 bp in length  
\* 18693 18792: gap of 100 bp  
\* 18793 19456: contig of 664 bp in length  
\* 19457 19556: gap of 100 bp  
\* 19557 20216: contig of 660 bp in length  
\* 20217 20316: gap of 100 bp  
\* 20317 21026: contig of 710 bp in length  
\* 21027 21126: gap of 100 bp  
\* 21127 21825: contig of 699 bp in length  
\* 21826 21925: gap of 100 bp  
\* 21926 22626: contig of 701 bp in length  
\* 22627 22726: gap of 100 bp  
\* 22727 23391: contig of 665 bp in length  
\* 23392 23491: gap of 100 bp  
\* 23492 24160: contig of 669 bp in length  
\* 24161 24260: gap of 100 bp  
\* 24261 24950: contig of 690 bp in length  
\* 24951 25050: gap of 100 bp  
\* 25051 25743: contig of 693 bp in length  
\* 25744 25843: gap of 100 bp  
\* 25844 26548: contig of 705 bp in length  
\* 26549 26648: gap of 100 bp  
\* 26649 27305: contig of 657 bp in length  
\* 27306 27405: gap of 100 bp

\* 27406 28101: contig of 696 bp in length  
\* 28102 28201: gap of 100 bp  
\* 28202 28913: contig of 712 bp in length  
\* 28914 29013: gap of 100 bp  
\* 29014 29714: contig of 701 bp in length  
\* 29715 29814: gap of 100 bp  
\* 29815 30502: contig of 688 bp in length  
\* 30503 30602: gap of 100 bp  
\* 30603 31249: contig of 647 bp in length  
\* 31250 31349: gap of 100 bp  
\* 31350 32023: contig of 674 bp in length  
\* 32024 32123: gap of 100 bp  
\* 32124 32821: contig of 698 bp in length  
\* 32822 32921: gap of 100 bp  
\* 32922 33592: contig of 671 bp in length  
\* 33593 33692: gap of 100 bp  
\* 33693 34381: contig of 689 bp in length  
\* 34382 34481: gap of 100 bp  
\* 34482 35163: contig of 682 bp in length  
\* 35164 35263: gap of 100 bp  
\* 35264 35949: contig of 686 bp in length  
\* 35950 36049: gap of 100 bp  
\* 36050 36745: contig of 696 bp in length  
\* 36746 36845: gap of 100 bp  
\* 36846 37544: contig of 699 bp in length  
\* 37545 37644: gap of 100 bp  
\* 37645 38336: contig of 692 bp in length  
\* 38337 38436: gap of 100 bp  
\* 38437 39103: contig of 667 bp in length  
\* 39104 39203: gap of 100 bp  
\* 39204 39867: contig of 664 bp in length  
\* 39868 39967: gap of 100 bp  
\* 39968 40624: contig of 657 bp in length  
\* 40625 40724: gap of 100 bp  
\* 40725 41403: contig of 679 bp in length  
\* 41404 41503: gap of 100 bp  
\* 41504 42191: contig of 688 bp in length  
\* 42192 42291: gap of 100 bp  
\* 42292 42955: contig of 664 bp in length  
\* 42956 43055: gap of 100 bp  
\* 43056 43724: contig of 669 bp in length  
\* 43725 43824: gap of 100 bp  
\* 43825 44449: contig of 625 bp in length  
\* 44450 44549: gap of 100 bp  
\* 44550 45257: contig of 708 bp in length  
\* 45258 45357: gap of 100 bp  
\* 45358 46045: contig of 688 bp in length  
\* 46046 46145: gap of 100 bp  
\* 46146 46838: contig of 693 bp in length  
\* 46839 46938: gap of 100 bp  
\* 46939 47617: contig of 679 bp in length  
\* 47618 47717: gap of 100 bp  
\* 47718 48404: contig of 687 bp in length  
\* 48405 48504: gap of 100 bp  
\* 48505 49183: contig of 679 bp in length  
\* 49184 49283: gap of 100 bp  
\* 49284 49968: contig of 685 bp in length  
\* 49969 50068: gap of 100 bp  
\* 50069 50714: contig of 646 bp in length  
\* 50715 50814: gap of 100 bp  
\* 50815 51333: contig of 519 bp in length  
\* 51334 51433: gap of 100 bp  
\* 51434 52144: contig of 711 bp in length  
\* 52145 52244: gap of 100 bp  
\* 52245 52932: contig of 688 bp in length  
\* 52933 53032: gap of 100 bp  
\* 53033 53710: contig of 678 bp in length  
\* 53711 53810: gap of 100 bp

Query Match 1.8%; Score 57; DB 2; Length 63173;  
Best Local Similarity 100.0%; Pred. No. 1.2e-19;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 333 aggggtattgagaatcaaccaagaagtttcataagccttcattcttcg 389
|||||
Db 61533 AGGGTGATTGAGATCAACCAAGAAGTTTCATGAAGCCCTCATCTCTCGG 61477

RESULT 12
AX015901
LOCUS
DEFINITION Sequence 3 from Patent WO9950284.
ACCESSION AX015901
VERSION AX015901.1 GI:10041643
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1140)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 3 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Source
1..1140
/organism="Homo sapiens"
/db_xref="taxon:9606"
<2..784
/feature="unnamed protein product"
/feature_start=1
/feature_id="CAC07619.1"
/db_xref="GI:10041644"
/translation="IHRAPPEQALLRRHPPQTRMLSDLVFCDOMGLPWFSSAGAL
NMALVFCGLLODPAOFRHVNLVPLTHFTSPIRRFADVLVHRLAALGYRERLDM
APDTLOQADHCDNRWASKRQVELSTLFFAVLVKESGPLESEAMVMGILKQAFDVL
VLRYGQKRLVCNALALRSHFKQVKKPELTLVPEDEMGEQPAQVITISLVEVY
LQAESTALKSAILKRPQTGHLGPEKEEESDGEPEDSSTS"
BASE COUNT 231 a 355 c 334 g 220 t
ORIGIN

Query Match 1.6%; Score 49; DB 6; Length 1140;
Best Local Similarity 100.0%; Pred. No. 4e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2815 tttttaattggttttaacaactcagggtttgtttttattttattt 2863
|||||
Db 866 TTTTAAATTGGTTTAAACAACCTCAGGGGTTTGTATTATTATTATT 914

RESULT 13
AX015910
LOCUS
DEFINITION Sequence 13 from Patent WO9950284.
ACCESSION AX015910
VERSION AX015910.1 GI:10041653
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1784)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 13 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Source
1..1784
/organism="Homo sapiens"
/db_xref="taxon:9606"
<3..1454
/feature="unnamed protein product"
/feature_start=1
/feature_id="CAC07621.1"
/db_xref="GI:10041654"
/translation="NFKVGVHIADVSFYFEGSDLDKVAARATSVYLIVKVVPMPLPR
LLCEELCSLNPSMDKLTFSVITLTPEKILDEWFGRTIIRSCTKLSYEHQAQSMIESP
TERIPAKELPPISPERSSEEVHQAVLNHLGIAKQLRQRFVDGALRLDQLKLAFTLDH
ETGLPQCHIEYERESNKLVEFMALHMAVAHIHRAFPQOALIRRRPPQTRMLSD
LVFCDOMGLPWFSSAGALNKLSQTQFGDKYSLARKEVLTNMCSPRMOMALYFCSG
LLODPAOFRHVNLVPLTHFTSPIRRFADVLVHRLAALGYRERLDMADPTLOQAD
HCDNRWASKRQVELSTLFFAVLVKESGPLESEAMVMGILKQAFDVLVIRYGVOKR
IYCNALALRSHFKQVKRPETLVPEDEMGEQPAQVITISLVEVVLQAESTALK
YSAILKRPQTGHLGPEKEEESDGEPEDSSTS"
BASE COUNT 385 a 537 c 504 g 358 t
ORIGIN

Query Match 1.6%; Score 49; DB 6; Length 1784;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2815 tttttaattggttttaacaactcagggtttgtttttattttattt 2863
|||||
Db 1536 TTTTAAATTGGTTTAAACAACCTCAGGGGTTTGTATTATTATTATT 1584

RESULT 14
AX015904
LOCUS
DEFINITION Sequence 7 from Patent WO9950284.
ACCESSION AX015904
VERSION AX015904.1 GI:10041647
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 13330)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 7 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Source
1..13330
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2787 a 4068 c 3977 g 2498 t
ORIGIN

Query Match 1.6%; Score 49; DB 6; Length 13330;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2815 tttttaattggttttaacaactcagggtttgtttttattttattt 2863
|||||
Db 6843 TTTTAAATTGGTTTAAACAACCTCAGGGGTTTGTATTATTATTATT 6891

RESULT 15
AX015914
LOCUS
DEFINITION Sequence 17 from Patent WO9950284.
ACCESSION AX015914
VERSION AX015914.1 GI:10041657
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 49999)
/feature="unnamed protein product"
/feature_start=1
/feature_id="CAC07621.1"
/db_xref="GI:10041654"
/translation="NFKVGVHIADVSFYFEGSDLDKVAARATSVYLIVKVVPMPLPR
LLCEELCSLNPSMDKLTFSVITLTPEKILDEWFGRTIIRSCTKLSYEHQAQSMIESP
TERIPAKELPPISPERSSEEVHQAVLNHLGIAKQLRQRFVDGALRLDQLKLAFTLDH
ETGLPQCHIEYERESNKLVEFMALHMAVAHIHRAFPQOALIRRRPPQTRMLSD
LVFCDOMGLPWFSSAGALNKLSQTQFGDKYSLARKEVLTNMCSPRMOMALYFCSG
LLODPAOFRHVNLVPLTHFTSPIRRFADVLVHRLAALGYRERLDMADPTLOQAD
HCDNRWASKRQVELSTLFFAVLVKESGPLESEAMVMGILKQAFDVLVIRYGVOKR
IYCNALALRSHFKQVKRPETLVPEDEMGEQPAQVITISLVEVVLQAESTALK
YSAILKRPQTGHLGPEKEEESDGEPEDSSTS"
BASE COUNT 385 a 537 c 504 g 358 t
ORIGIN
```

```
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC07621.1"
/db_xref="GI:10041654"
/translation="NFKVGVHIADVSFYFEGSDLDKVAARATSVYLIVKVVPMPLPR
LLCEELCSLNPSMDKLTFSVITLTPEKILDEWFGRTIIRSCTKLSYEHQAQSMIESP
TERIPAKELPPISPERSSEEVHQAVLNHLGIAKQLRQRFVDGALRLDQLKLAFTLDH
ETGLPQCHIEYERESNKLVEFMALHMAVAHIHRAFPQOALIRRRPPQTRMLSD
LVFCDOMGLPWFSSAGALNKLSQTQFGDKYSLARKEVLTNMCSPRMOMALYFCSG
LLODPAOFRHVNLVPLTHFTSPIRRFADVLVHRLAALGYRERLDMADPTLOQAD
HCDNRWASKRQVELSTLFFAVLVKESGPLESEAMVMGILKQAFDVLVIRYGVOKR
IYCNALALRSHFKQVKRPETLVPEDEMGEQPAQVITISLVEVVLQAESTALK
YSAILKRPQTGHLGPEKEEESDGEPEDSSTS"
BASE COUNT 385 a 537 c 504 g 358 t
ORIGIN

Query Match 1.6%; Score 49; DB 6; Length 1784;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2815 tttttaattggttttaacaactcagggtttgtttttattttattt 2863
|||||
Db 1536 TTTTAAATTGGTTTAAACAACCTCAGGGGTTTGTATTATTATTATT 1584

RESULT 14
AX015904
LOCUS
DEFINITION Sequence 7 from Patent WO9950284.
ACCESSION AX015904
VERSION AX015904.1 GI:10041647
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 13330)
AUTHORS Aigner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 7 07-OCT-1999;
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
Source
1..13330
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2787 a 4068 c 3977 g 2498 t
ORIGIN

Query Match 1.6%; Score 49; DB 6; Length 13330;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2815 tttttaattggttttaacaactcagggtttgtttttattttattt 2863
|||||
Db 6843 TTTTAAATTGGTTTAAACAACCTCAGGGGTTTGTATTATTATTATT 6891

RESULT 15
AX015914
LOCUS
DEFINITION Sequence 17 from Patent WO9950284.
ACCESSION AX015914
VERSION AX015914.1 GI:10041657
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 49999)
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 07:57:53 ; Search time 3767.8 Seconds  
(without alignments)  
17217.544 Million cell updates/sec

Title: US-09-647-377-8  
Perfect score: 3100  
Sequence: 1 cggcgccgcgcgcctccogg.....tgaaaaaaaaaaaaaaaaaa 3100

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_ov.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	3100	100.0	3100	6	AX015905	AX015905 Sequence
2	1330.6	42.9	1550	6	AX015899	AX015899 Sequence
3	1181.8	38.1	1784	6	AX015910	AX015910 Sequence
4	565.6	18.2	1140	9	AB066536	AB066536 Macaca fa
5	558.4	18.0	2189	9	AB066536	AB066536 Macaca fa
6	473	15.3	36901	6	AX015903	AX015903 Sequence
7	473	15.3	38886	6	AX015909	AX015909 Sequence
8	473	15.3	228283	10	AF312994	AF312994 Mus muscu
9	247.4	8.0	82454	8	AC012193	AC012193 Arabidops
10	236.8	7.6	49999	6	AX015902	AX015902 Sequence
11	236.8	7.6	49999	6	AX015908	AX015908 Sequence
12	230	7.4	3597	3	AY058669	AY058669 Drosophil
13	222	7.2	153701	2	AC098189	AC098189 Rattus no
14	215.4	6.9	162842	2	AC024062	AC024062 Homo sapi
15	215.4	6.9	345102	9	AF312225	AF312225 Homo sapi
16	197.8	6.4	35000	8	SPAC2C4	Z99259 S.pombe chr
17	187.6	6.1	156905	2	AC020006	AC020006 Drosophil
18	187.6	6.1	199016	3	AC008204	AC008204 Drosophil
19	187.6	6.1	204442	3	AE003747	AE003747 Drosophil
20	186.6	6.0	132635	2	AC007577	AC007577 Drosophil
21	183.4	5.9	131853	2	AC017755	AC017755 Drosophil
22	183.4	5.9	153701	2	AC098189	AC098189 Rattus no
23	183.4	5.9	168250	3	AC093121	AC093121 Drosophil
24	183.4	5.9	173494	3	AC093501	AC093501 Drosophil
25	183.4	5.9	174279	2	AC006169	AC006169 Drosophil
26	183.4	5.9	190351	3	AC105293	AC105293 Drosophil
27	183.4	5.9	298640	3	AE003467	AE003467 Drosophil
28	182.4	5.9	3129	3	AY052150	AY052150 Drosophil
29	182.2	5.9	13330	6	AX015904	AX015904 Sequence
30	182.2	5.9	49999	6	AX015914	AX015914 Sequence
31	182.2	5.9	175391	2	AC093374	AC093374 Homo sapi
32	182.2	5.9	181068	9	AC068134	AC068134 Homo sapi
33	179.4	5.8	35325	3	U23514	U23514 Caenorhabdi
34	176	5.7	49999	6	AX015912	AX015912 Sequence
35	159.8	5.2	3602	3	DME276897	AJ276897 Arabidops
36	156	5.0	2992	8	AY057479	AY057479 Arabidops
37	150.4	4.9	49999	6	AX015913	AX015913 Sequence
38	147.8	4.8	162842	2	AC024062	AC024062 Homo sapi
39	147.4	4.8	3344	3	TBR308998	AJ308998 Trypanoso
40	145.2	4.7	139049	9	AC019130	AC019130 Homo sapi
41	145.2	4.7	219775	2	AC105461	AC105461 Homo sapi
42	138.4	4.5	2466	9	BC014124	BC014124 Homo sapi
43	138.4	4.5	3694	9	BC022089	BC022089 Homo sapi
44	137	4.4	219775	2	AC105461	AC105461 Homo sapi
45	131.6	4.2	4630	3	U00057	U00057 Caenorhabdi

ALIGNMENTS

RESULT	1	AX015905	Sequence 8 from Patent WO9950284.	3100 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX015905	Sequence 8 from Patent WO9950284.					
DEFINITION	AX015905	Sequence 8 from Patent WO9950284.					
ACCESSION	AX015905	Sequence 8 from Patent WO9950284.					
VERSION	AX015905.1	GI:10041648					
KEYWORDS							
SOURCE		house mouse.					
ORGANISM		Mus musculus					
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.					
REFERENCE		Aligner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.					
AUTHORS		Aligner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.					
TITLE		Nucleic acid molecules which code proteins influencing bone development					
JOURNAL		Patent: WO 9950284-A 8 07-OCT-1999;					
		ALIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP					
		ANDREAS (DE); WIRTH THOMAS (DE)					
FEATURES		Location/Qualifiers					
source		1..3100					
		/organism="Mus musculus"					
		/db_xref="taxon:10090"					
		125..2737					
CDS							

```
/note="unnamed protein product"  
/codon_start=1  
/protein_id="CAC07620.1"  
/db_xref="GI:10041649"
```

```
translation="MNHDPYKLNLRSPGTGRVSVVGPSSAVGASPDGRKSKNKSMMRG  
KKKSIFFETYSKEDVEBGLRGTLIOGVLRINPKKFHEAFIPSPDGRDIFIDGVVAR  
NRALNGOLVWVKLLPDEDMKAVKPESNDKETIATYADIPEEGGHPHLOQSRKWSG  
PDLVIEAQFDDSDSDRHGNTSLVDGKLSISTPDRKEDSSTPMKDNTPILPOD  
TRGLSEKLSOKSAKVVILEKKHSRAATGILKLLADKNSDLFKKYALESPSHRVPRI  
VYPLKDCPQPMTRPKDFANTLFIGRIIDWKEDCNFALQOLAKLSLQAGAEIEPFEFIGI  
LTEYGVDSFESSEVLKLECLPQSLPWTIIPDEVGRRLRKDCIFITDIPSTADLDLAL  
ACRLTIDGTFEVGHADIAVSVFVPESSLDVAAERATSVYLKQVYMLPRLICEEL  
CSLPMTDIKLTF SVTWKLTPEGLKLSIAEKFQRIIRRSCTKLSYDHAQSMTENETKPIE  
BELPPIISPEHSVEEVHQAIVLNHLSIAKORRFVDGALRLDLKLAFTLDHETLPO  
GCHTIEYRDSNKLVEEFMLLANNAVAHKIFRTFPEQALLRRHPPOTKMLSDLVFRC  
QWGLPMDVSSAGALNKLTKTFGDDKYSLARKEVLTNNYSRPMOMALYPCSGMLDDOE  
OFRIYALNPLYTHFTSPIRRFADVILHRLAAALGYSEOPDEVPETLOKQADHNDR  
EMASKRVQELSIGLFFAVLVKESGPLESEFMVMGVLNQADFVLVLRFGVQKRIYCNAL  
ALRSYSQKVGKKPELTLVWEPDDLEEBPTQOVIITIFSLVDVVQLQEAATALYSAILK  
RPLGEKASDEEPED"
```

BASE COUNT 783 a 795 c 812 g 710 t  
ORIGIN

Query Match 100.0%; Score 3100; DB 6; Length 3100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 cggcgccggcgctccggagagcagctcgtagaacatgagctgtagaagcgaggg 60  
Db 1 CGGCGCCGGCGCTCCGGAGGCGGCTCGTGAACATGAGCTGAGCTGAAGCGAGG 60  
QY 61 aactctgagctgaatagtagtgggtccctgaatctggagagaagagcgcaacttgaacc 120  
Db 61 AACTCTGAGCTGAATAGTAGTGGTCCCTGAATCTGAGAGAGAGCGCACCTTGAACC 120  
QY 121 agtaataaacatcctgaactaagctgaaccttgggtcccgaggagaccccgaggagt 180  
Db 121 AGTAATGAACCATCTGACTACAAGCTGAACCTTCGGTTCGCGGGAGCCCCAGAGGTG 180  
QY 181 gtccctgtggtggccggagtcgtgtggtgcttcgccaggtgacaaaaagtcacaagaa 240  
Db 181 GTCCCTGTGTGGTGGCCGGAGTGTGTGTGGTGTTCGCCAGGTGACAAAAAGTCAAGAA 240  
QY 241 caagtcctatcgaggaagaaagagacatatatttaaacctacatgtccagaggagagt 300  
Db 241 CRASTCCATGCGAGGGAAGAAAAGACATATTTGAACCTACATGTCCAGAGGAGGATGT 300  
QY 301 ttccagaaggttgagagagaacacattatccagggtgtattgagaaatcaacccaagaa 360  
Db 301 TTCAGAAGGCTTTGAAGAGAGAGACACTTATCCAGGCTGTATTGAGAATCAACCCAAGAA 360  
QY 361 gttccatgaagccttcattcctctccgagtgatcgagacattttttattatgaagt 420  
Db 361 GTTTCATGAAGCCTTCATCTCTTCGCGATGGTGTATCGGACATATTTTATTGATGAGT 420  
QY 421 tgttctcgttaatagagccttaaatgggacaccttgggtgtgtaaaactgtcctcagaga 480  
Db 421 TGTTCGTCGTAATAGAGCCTTAAATGGGACCTTGTGTGTGTGTAATAACTGTCTCTGAGGA 480  
QY 481 tcaagtgaagggcagttaaaccagagagcaatgacaaaagaataaagctacttatgaagc 540  
Db 481 TCAGTGGAGGCGAGTTAAACCAGAGAGCAATGACAAAGAAATAGAAGCTACTTATGAAGC 540  
QY 541 tgacatccctgaaggggctgtggacatcacccctgcagcagtcctcggaagggctggag 600  
Db 541 TGACATCCCTGAAGAGGCGCTGTGACATCACCCCTCGCAGAGTCCCGGAAAGGCTGGAG 600  
QY 601 tggctctgatgtcattatagagctcagttttatcacagcaactcaaatagacatagac 660  
Db 601 TGGTCTGTATGTCATTATAGAGCTCAGTTTGTATGACAGGACACTCAGAAATAGACATG 660  
QY 661 caacaccagtggtgtgtatgaggtgttaagaaattgtcaatctcctcctgacagagg 720
```

```
Db 661 CAACACCAAGTGGCTGGTTGATGGTGAAGAAATTTGCAATCTCTACTCTGACAGAGG 720  
QY 721 aaaaagaattcttagtactccagttatgaagatgaagaacacccccataccacagagcac 780  
Db 721 AAAAGAAGATTCTAGTACTCCAGTTATGAAGATCAGACACACCCCATACCACAGACAC 780  
QY 781 aagaagcttatcagaagaagtcactcaagaatacagcaaaaagtggtttacatcttgagaa 840  
Db 781 AAGAGGCTTATCAGAGAAGTCACTTCAGAAATCAGCAAAAGTGGTTTACATCTTCGAGAA 840  
QY 841 aaagcattctcagcagcaactggatcctctgaacctcttgctgataaagaacagtgacct 900  
Db 841 AAAGCATTTCTCGACAGCAACTGGCATCTCTGAAACTCTTGGCTGATAAGAACAGTACCT 900  
QY 901 gtttaagaatacgcctgttttctccttcagacacacccagtagtacccttagaatttaagta 960  
Db 901 GTTTAAGAAATACGCGCTGTTTCTCTTCAGACACCGAGTACCTTAGAATTTAGGTACC 960  
QY 961 tctcaagactgtcccaagcagctcatgaccgcacccctaaagactttgccaacacgctgtt 1020  
Db 961 TCTCAAGGACTGTCCCGAGGACTTCATGACCCGACCTTAAAGACTTTGCCAACACAGCTGT 1020  
QY 1021 catctgcgcacatagactgaagagagactgtaatttgcctgggggaactgctaa 1080  
Db 1021 CATCTGCGCATCATAGACTGAAGAGGAGCTGTAAATTTGCCCTGGGCAACTGGCTAA 1080  
QY 1081 gagtctggcgagctgggtgaactcgagcctgaacagagaaggaactactgacagaatatgg 1140  
Db 1081 GAGTCTTTGGCGAGCTGGTGAATCGAGCTTGAACAGAGAGGATACTGACAGAATATGG 1140  
QY 1141 tggagactctctgattctctcagaagttcttgaatgtctccctcaagcgtgcctg 1200  
Db 1141 TGTGGACTTCTTGATTTCTCTTCAGAAGTTCTTGAATGTCTCCCTCAAGGCTGCCCTG 1200  
QY 1201 gacaaatccacctgatgaggtgggcaagagaagacctaagaagaagactgtattctaac 1260  
Db 1201 GACAATCCCACTGATGAGGTGGGCAAGAGAAGAGACCTAAGGAAAAGACTGTATCTTCAC 1260  
QY 1261 cattgattccatcaactgcgcgaccttggatgctcctgcctcagcggtcactga 1320  
Db 1261 CATTGATCCATCAACTGCTCGAGCCTTGTGATGATGCGCTCGCTCGAGGGGCTCACTGA 1320  
QY 1321 tggcaccttcgaagtggggcgctccacaatcgccgagtgtgagttacttctctgagggatc 1380  
Db 1321 TGGCACCTTCGAAGTGGGCGTCCACATCGCGATGTGAGTTACTTTGTTCCTGAGGGATC 1380  
QY 1381 ctctttggataaagtagctgtgagagagccacaagtgtctacttgggtccagaaggtgg 1440  
Db 1381 CTCCTTTGGATAAAGTAGCTGTGAGAGAGCCACAAGTGTCTACTTGGTCCAGAGGTGGT 1440  
QY 1441 ccccatcttccaggctctctgttgaggaactcgcagcctcaacccccatgactgacaa 1500  
Db 1441 CCCCATCTTCCAGGCTTCTGTGTGAGGAACCTGTGAGCCTCAACCCCATGACTGACAA 1500  
QY 1501 gctgacctctctgtgactgtgaaactgaacccctgaaggcaagatccttgaagagtggt 1560  
Db 1501 GCTCACCTTCTCTGTGATCTGGAAGCTGACCCCTGAAGSCAAGATCCTTGAAGAGTGGT 1560  
QY 1561 tggccgcactataatccgtttctgaccacaactgagcagacatgcccagagcatgat 1620  
Db 1561 TGGCCGCACTATCATCCGTTCTTGACCACTAAGCTGAGCTACGACCATGCCAGAGCATGAT 1620  
QY 1621 cgaataatcccaactgagaagatccctgaggaagagcttcccccaatttctccagagacag 1680  
Db 1621 CGAAATCCCAACTGAGAAGATCCCTGAGGAGAGCTTCCCCCAATTTCTCCAGAGCACAG 1680  
QY 1681 cgtcgaagaggtgcaccaggaagtcctctgaacctgacagcatgtcaagaagcaactccg 1740  
Db 1681 CGTCGAGGAGGTGCACAGGAGAGTCTCTGAACCTGCACAGCATTGCAGAAAGCAACTCC 1740  
QY 1741 ccagcgtcttggtagaaggcgcaactccgttttagatcagctgaagcgtgttttactctgga 1800
```

Db	1741	CCAGCGCTTTGTAGATGGCGCACTCCGTTTGTAGATCAGCTGAAGCTTGCTTTTACTCTGGA	1800
Qy	1801	ccatgagactggaactgcctcaagatgtcacaatctatgagtaccagagacaagaact	1860
Db	1801	CCATGAGACTGGACTGCCCTCAAGGATGTCACATCTATGATGATCCGAGAGACAGCAAGAAGCT	1860
Qy	1861	tgtagaggatctcaactcctcctggccaacatggcggtggcccaagatcttcccaactt	1920
Db	1861	TGTAGAGAGGTTCTGCTCTGTGGCCACATGGCGGTGGCCACAGATCTTCGCGACCTT	1920
Qy	1921	ccctgagcagccctgctgcgcgcgcatcccccaaccacagaagaagtctcaagtgaact	1980
Db	1921	CCCTGAGCAGGCGCTGCTCGCGCGCATCCGCCACACAGACGAAGATGCTCAGTGAAGCT	1980
Qy	1981	ggtgaactctgtgaccaagtggcctgccatggaagtcaagctctgaggggcccctaaa	2040
Db	1981	GGTGAGGTTCTGTGACCATGGGCTGCCCATGGATGTCAGCTCTGCAGGGGCCCTTAA	2040
Qy	2041	taaaagcctgaactaagacattgggagatgacaagtactctctgcccggaaagagtgct	2100
Db	2041	TAAAAGGCTGACTAAGACATTTGGAGATGACAAGTACTCTGTGCCCGGAAGAGGTGCT	2100
Qy	2101	caccaactatctcccgcccatcagatggcaactgtactctgtctgggatctgca	2160
Db	2101	CACCAACATGTCTCCGCGCCATGCAGATGGCACTGTATTCTGTCTGGGATCTGTCA	2160
Qy	2161	ggaccagagcagttccggcattatgtctcacaagttccctctcacacacttcaactc	2220
Db	2161	GGACCAGAGCAGTTCGGGCATATATGCTCTCAAGGTTCCCTCTACACACACTTCACCTC	2220
Qy	2221	tcctatccgcgcttctgtagcgtcatagtacccgctcctgctgctgtctgtggcta	2280
Db	2221	TCCATTCGCGCTTCTGTGAGCTCATAGTCACCGGCTCCTGGCTGTCTGGGCTA	2280
Qy	2281	cagtgaaacagcagatgtgagcctgtatccctcacagaagaagcagctgaccactgca	2340
Db	2281	CAGTGAACAGCCAGATGGAGCCTGATACCTACAGAAACAAAGCTGACCACTGCAATGA	2340
Qy	2341	ccgtgcagcttccaaacgtgtgcagagctcagatcgccctctcttcgcagttct	2400
Db	2341	CCGTGCATGGCTTCCAAAGCTGTGCAGAGCTCAGAGCTCAGCATCGGCTCTCTTCGCAGTTCT	2400
Qy	2401	agtaaaagagagtggtgccccctggagtcogaagccaatggtgatgggtgctcctgaaccaagc	2460
Db	2401	AGTAAAGAGAGTGGCCCCCTGGAGTCCGAAGCCATGGTGATGGGTGCTCTGAAACCAAGC	2460
Qy	2461	tttcgactgtctgtctgcgcttggggtgcagaagcgcatctactgcaatgcactggc	2520
Db	2461	TTTCGACGTGCTGTGCTGGCCCTTGGGCTGGCAGAGCGCATCTACTGCAATGCAGTGGC	2520
Qy	2521	cctgcgatactacagcttccagaagtggtgggaagcagcagagctcactctgtttggga	2580
Db	2521	CCTGCGATCCTTACAGCTTCCAGAGGTGGGGAAGAACGACAGAGCTCATCTCTTTTGGGA	2580
Qy	2581	gctgaatgacctaaagagagccaacacagcagctcatcaccactctcagcctgtgga	2640
Db	2581	GCCTGATGACCTTGAAGAGAGGCCAACACAGCAGCTCATCACCATCTTCAGCCTGTGGA	2640
Qy	2641	tgtggtcctgcaggagagccacagccctcaagtcaagtcactctcctgaagcaccagg	2700
Db	2641	TGTGTGCTGTGAGGCGAGAGGCCACAGCCCTCAAGTACAGTACTCTTCTGAAAGGACAGG	2700
Qy	2701	cctggagaagcgctctgatgagagcctgagactgaatgctagcccaagccagcgctgt	2760
Db	2701	CCTGGAGAAGCGCTCTGATGGAGGCGCTGAGGACTGAATGCTAGCCCAAGCCAGGCTGT	2760
Qy	2761	gcctgcccatacctgtcgtgttttagaataaggaccttttagacacaaagggattttta	2820
Db	2761	GCCTGCCCTACCTGCTGGCTTTTAGAATAGAGACCTTTTGACACAAAGGGGATTTTA	2820
Qy	2821	atttggttttaaacactcaagggtttgtttttattttttatttttttttttttttactt	2880
Db	2821	ATTTGGTTTTTAACAACTCAGGGTTTGTTTTATTTATTTTTCCTTTTATTTTACTT	2880
Qy	2881	ttgcagctcagtttttaataactgaagggttaagggtcagggcagggagtgctgaggc	2940
Db	2881	TTGCAGCTCAGTTTTTAAATGAACCTGAAGGTTAGGGTCAAGGAGGGAGGTGCTGAGGC	2940
Qy	2941	ctggcctgtgtctcctgagcagagagatccctcctcctcctcctcctcctcctcctcctt	3000
Db	2941	CTGGCCTGTGCTTCCCTCAGCAGAGAGGATCCAGCTCCTCTGGCGAGCGCCCGCTT	3000
Qy	3001	ctaccagcagccactgccttccctgccaggaatgggggttttcagcaaatcagtg	3060
Db	3001	CTACCAAGCGACCACTGCCCTTCCCTGCCAGGAATGGGGGCTTTCAGCAAAATCAGTG	3060
Qy	3061	tcattgaataaaatcaagtgtaaaaaaataaaaaaataaaaaa 3100	
Db	3061	TCATGGAATAAAATCAAGTGTGAAAAAATAAAAAAATAAAAAA 3100	
RESULT 2			
AX015899			
LOCUS			
Sequence 1 from Patent WO9950284.			
DEFINITION			
AX015899			
ACCESSION			
AX015899.1 GI:10041641			
VERSION			
KEYWORDS			
SOURCE			
house mouse.			
ORGANISM			
Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE			
1 (bases 1 to 1550)			
AUTHORS			
Aligner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.			
TITLE			
Nucleic acid molecules which code for proteins influencing bone			
development			
JOURNAL			
Patent: WO 9950284-A 1 07-OCT-1999;			
ALIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP			
ANDREAS (DE); WIRTH THOMAS (DE)			
FEATURES			
source			
1..1550			
/organism="Mus musculus"			
/db_xref="taxon:10090"			
<2..1183			
/note="unnamed protein product"			
/codon_start=1			
/protein_id="CAC07618.1"			
/db_xref="GI:10041642"			
/translation="LGRSKVAERATSVLVQKVPMLPRLLCEELCSLNPMTDKLTF			
SVYWKLTPECKILEWFGRTIIRSCIKLSYDHAQSMIENETKIPPEELPPISPESHV			
EEVHQAVLNHLSIAKQRRQRFVDGALRDLQEFMLLANMAVAHKIERTFPEQALLRRH			
PPQTKMLSDLVFECDQGLPMDVSVSAGALNMALYFCGMLQDEQFRHYALNPLVT			
HFTSPIRFADVTVHRLAAALGYSEQPDVPTLQKADHCNDRMRASKRYQELSIG			
LFFAVLVKESGPLESFAMVYGLNQAFDVLVLRFGVKRYICNALALRSYSFKQVKK			
PELILVWEPDDLEEEPTQOVITITFSLVDVLQAEATALKYSAILKRPGLKASDDEPE			
D"			
BASE COUNT			
352 a 433 c 412 g 353 t			
ORIGIN			
Query Match			
Best Local Similarity 42.9%; Score 1330.6; DB 6; Length 1550;			
Matches 1530; Conservative 0; Mismatches 4; Indels 177; Gaps 2;			
Qy	1390	taaaagtagctgtgagagagcacaagtgctacttggtccagaaggtggtcccaactt	1449
Db	13	TAAAGTAGCTGCTGAGAGAGCCAAAGTGTCTACTTGGTCCAGAGGTGGTCCCATGCT	72
Qy	1450	tcaccagctctctgtgagaaactctgcagctcaacccactgactgacaagctgaactt	1509
Db	73	TCCCAGGCTTCTGTGTGAGGAACCTCTGCAGCCTCAACCCCATGACTGACAGCTT	132
Qy	1510	ctctgtatctgaaactgaagcaccctgaagcgaagatacttgagagtggtttggccac	1569
Db	133	CTCTGTATCTGGAAGCTGACCCCTGAAGGCAAGATCTCTTGAAGAGTGGTTTGGCCGAC	192

QY 1570 tatcatcgtcttgcacaaactgactacgaccatgcccagagcatgatcgaaaaatcc 1629  
DB 193 TATCATCGTCTTGTGACCAAACTGAGCTACGACATGCGCAGAGCATGATCGAAATCC 252  
QY 1630 aactgagaactccctgaggaagacttcccacaaattctccagagacacagctcgaaga 1689  
DB 253 AACTGAGAAGATCCCTGAGGAAGAGCTTCCCCCAATTTCTCCNAGACACAGCTCGAGGA 312  
QY 1690 ggtgcaccaggcagctcctgaaacctgcacagcattgcaaaagcaactccgcgcagcgctt 1749  
DB 313 GGTGCACACAGCGCAGTCTCTGAACCTGCACAGCATTTGCAAAAGCAACTCCGCCCGCAGCGCTT 372  
QY 1750 tctagatggcgccactccgttttagatcagctgaagcttgctttactctcggaccatgagac 1809  
DB 373 TGTAGTGGCGCACTCCGTTTAGATC----- 398  
QY 1810 tggactgacctcaaggatgtcaatctatgagtaaccgagacagaacaagctgttagagga 1869  
DB 399 -----AGGA 402  
QY 1870 gttcatgctcctggccaaacatggcggtggcccaagaatcttcgcgaacctccctcgaqa 1929  
DB 403 GTTCATGCTCTGGCCAAACATGGCGGTGGCCCAACAAGATCTTCGCGACCTTCCCTGAGCA 462  
QY 1930 ggcctgctgcgcgcgcatcccccacacagaagaatgctcaagtacactggtagatt 1989  
DB 463 GGCCTGCTGCGCGCGCATCCCCCAACACAGCAAGATGCTCAGTGAACCTGGTGGAGATT 522  
QY 1990 ctgtgaccagatgggctgcccattggtcagctctgcaaggcgccctaaataaagcct 2049  
DB 523 CTGTGACAGATGGGCTGCCCATGGATGTCAGCTCTGCAAGGGGCCCTAAAT----- 574  
QY 2050 gactaagaacattggagatgacaagtactctctgcccggaggaggtgctcacaacaat 2109  
DB 575 ----- 574  
QY 2110 gtaactccggcccatcagatggcaactgtactctgctcgggaatgctgcaggaccagga 2169  
DB 575 -----ATGGCACTGTACTTCTGCTCTGGGATGCTGCAGGACCAAGA 615  
QY 2170 gcagttccggcattatgctctcaacgttcccccctacacacactcaactcccatccg 2229  
DB 616 GCAGTTCGGCATATATGCTCTCAAGGTTCCCTCTACACACACTTCACTCTCCCATCCG 675  
QY 2230 ccgcttgcagcgtcatatgacacccgctcctgctgctgctcgtgagctcaatgaaca 2289  
DB 676 CCGCTTGTGACGTCATATGTCACCGCTCTCTGGCTGCTGCTCTGGGCTACAGTGAACA 735  
QY 2290 gccagatgtggagcctgtatccctacagaagaagcgtgaccactgcaatgacctcgcat 2349  
DB 736 GCCAGATGTGGAGCCTGTATACCTTACAGAAGCAAGCTGACCACATGCAATGACCGTCGCAT 795  
QY 2350 ggcttccaaacgtgtcaggagctcagcatcgccctctcttcgcagttcttagtaaaagga 2409  
DB 796 GGCTTCCNAACGCTGCGAGAGCTCAGCATCGGCTCTTCTTCGCAAGTCTAGTAAAGGA 855  
QY 2410 gsgtggccctggagtccgaagccatggtgatgggtgctgctgaacaaagtcttcgacgt 2469  
DB 856 GAGTGGCCCCCTGGAGTCCCAAGCCATGGTATGGGTGTCTCTGAACCAAGCTTTTCGACGT 915  
QY 2470 gctggtgctgcgcttgggggtgcagaagcgcatactgcaatgcaactggccctgcgac 2529  
DB 916 GCTGGTCTCGGCTTTGGGGTGCAGAAGCATCTACTGCAATGCACTGGCCCTGCATC 975  
QY 2530 ctacagcttccagaaggtggggaagaagccagagctcaactctctgtttggagcctgatga 2589  
DB 976 CTACAGTCTCCAGAAAGTGGGGAAGAACCCAGAGCTACACTCTTGTGTTGGAGCGCTGATGA 1035  
QY 2590 ccttgaagaggagcacaacagcaggtcatcacactcttcagcctggtggaatggtgctcct 2649  
DB 1036 CTTTGAAGAGAGGCCAACACAGCAGGTCATCACCATCTTCAGGCTTGGTGGATGTTGCTCT 1095  
QY 2650 gcaggcagaggccacagccctcaagtacagtgtctatctgtaagcagaccagccctggagaa 2709

DB 1096 CGAGGCAGAGGCCACAGCCCTCAAGTACAGTGCCTATCTTGAAGCGACCCAGGCTTGAGAA 1155  
QY 2710 ggcctctgtagagagcctgagagactgaatgctagcccagccagggcctgtgctgcctc 2769  
DB 1156 GCGCTCTGATGAGAGGCCCTGAGGACTGAATGCTAGCCCAAGCCAGGCTGTGCTGCCCT 1215  
QY 2770 accctgctgcttttaggaataagacaccttttgacacaaaggggatttttaatttggttt 2829  
DB 1216 ACCCTGCTGCTTTTAGGAATAGGACCTTTTGACACAAAGGGGATTTTAAATTTGGTTT 1275  
QY 2830 ttaacaactcaggggttcttttattttatttttcttttatttttatttttacttttgcagctc 2889  
DB 1276 TTAACAACCTCAGGGGTTCTTTTTATTATTTATTTTCTTTTATTATTTTGTGACGTC 1335  
QY 2890 agtttttaaatgaactggaaggttaggggtcagggcaggggagctgctgagcctggcctgt 2949  
DB 1336 AGTTTTAAATGAACGTGAAGGTTAGGGGTACGGGAGGGGATGCTGAGGCTGGGCTGT 1395  
QY 2950 gcttcctgagcagagagatcccagtcctcctgggcagcgagcccccctcttaccaggc 3009  
DB 1396 GCTTCCCTGAGCAGAGAGATCCCAGTCTCTCTGGCAGCAGCCCGCTTCTTACCAGGC 1455  
QY 3010 gaccactgctcctccctccaggaataagggggtttcagcaaatcagttcatgaaat 3069  
DB 1456 GACCATCTGCCCTTCCCTCCCGGAATGGGGGTTTCAGCAATCATGTCTATGGAAT 1515  
QY 3070 aaaaatcaagtgtgaaaaaataaaaaa 3100  
DB 1516 AAAATCAAGTGTGAATTGCAAAAAA 1546  
RESULT 3  
AX015910  
LOCUS AX015910 1784 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 13 from Patent WO9950284.  
ACCESSION AX015910  
VERSION AX015910.1 GI:10041653  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 1784)  
Aigner, T., Hess, J., Rosenthal, A., Rump, A. and Wirth, T.  
TITLE Nucleic acid molecules which code proteins influencing bone development  
JOURNAL Patent: WO 9950284-A 13 07-OCT-1999;  
AIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP ANDREAS (DE); WIRTH THOMAS (DE)  
FEATURES  
source 1. .1784  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07621.1"  
/db\_xref="GI:10041654"  
/translation="NEKVGVIADVSFVPEGSDLDKVAERATSYLVQKVPMLPR  
LLCEELCSLNPMSDKLTFVITLTPEGKILDEWFGRTIIRSCTKLSVEHAQSMIESP  
TBKIPAKELPPISEHSSERVHQAIVLNLGHAKOLROOFVDGALRDOLKLAFTLDH  
ETGLPOGCHIVEYBESNKIVEPMILANVAHKITHAFPEQALLRHPPTPMLSD  
LVFECQMGFLPVDFFSSAGALNLSLTQTGDDKYSILARKEVLTNCSRPMQMALFYCSG  
LIDQPAQRHYALNVPLTYHTFSPIRRFADVLVHRLAALAGILKQAFDLVLRYGVQR  
DCHDNRRWAKRVOELSTSLFFAVLVKSGSEPMVMGILKQAFDLVLRYGVQR  
IYCNALRSHHFQKVKPELTLVWPEDEMQEPAQQVITIFSLVEVVLQAEALTA  
YSAILKRPQTGCHLGPKEEESDGEEDSSTS"

Query Match 38.1%; Score 1181.8; DB 6; Length 1784;

Best Local Similarity 81.9%; Pred. No. 2.le-296; Matches 1471; Conservative 0; Mismatches 267; Indels 59; Gaps 7;			
QY	1323	gcaccttcgaagtgggcgtccacatcgccgatgtgagttactttgtctcctgagggatcct	1382
Db	1	GCAACTTCAAAAGTGGGAGTTACATTGCTGACGTGAGTTACTTTGTTCCGGAGGATCTG	60
QY	1393	ctttggataaagttagctgtgagagaccacaagtgtctacttggctccagaagtggtcc	1442
Db	61	ATCTGGATAAAGTGGCTGCCGAGAGGGCTACAAGCGTCTACTTTGGTTCAAAAGTGGTCC	120
QY	1443	ccatgtctccaggtctctgtgtgaggaactctgacgtccaccccatgactgacagc	1502
Db	121	CCATGTTCCCAGGCTGCTGTGTGAGGAGCTGTGCAGSCCTCAACCCATGTCCGACAAAG	180
QY	1503	tgaacctctctgtgactctggaagctgaacctgaagcgaagatcctttgaagtagtggttg	1562
Db	181	TGACCTTCTCTGTGATCTGGACACTGACTCCAGAGGGCAAGATCTCTTGATGAATGGTTTG	240
QY	1563	gccgcactatcctcgtttctgcaccaaactgactacgaccatgcccagagcatgatcg	1622
Db	241	GCCGGACCATCATCCGCTCCTGCACCAAACTTAGCTACGAGCATGCACAGAGCATGATTG	300
QY	1623	aaaatccaaactgagaagatccctgaggaagcttccccaatttctccagagcacagc	1682
Db	301	AAAGCCCAACTGAGAAATTCCTCGAAAGAGCTGCCCCCAATTTCCCCAGAGCATAGCA	360
QY	1683	tcgaggagggtgcaccaggcagtcctgaacctgcacagcattgcaaaagcaactccgcgcc	1742
Db	361	GCGAGGAGGTACACACAGGCGCTTGAATCTCCACGGAATGCCAAGCAGTTACGCCAGC	420
QY	1743	agccctttgtagatggcgcaactccgttagatcagctggaagctgttttactctggacc	1802
Db	421	ACGCGTTTGTGGACGGGCACCTCGTTTGGATCAGCTAAAGCTGTGTTTCACTCTGGACC	480
QY	1803	atgagactggaactgctcgaagatgtcacatctatgtagtaccagagacacaagottg	1862
Db	481	ACGAGACCGGATTTGCCCTCAAGGATGTCTATATATGATGATGATACCGAGACCAAGCTCG	540
QY	1863	tagagaggttcactccttggccaacatggcggtagcccaagatcttccgcacttcc	1922
Db	541	TGGAGGAGTTTACGTCTCTTGGCCACATGGCAGTGGGCCACAGATCCACCGGCGCTTCC	600
QY	1923	ctgagcaggccctgtcgcgcggcatcccccaccacagacgaagatgctcagtgacctgg	1982
Db	601	CCGACGAGGCCCTGTGCGCGCGCACCCCGCCCAAAAGATGCTCAGTGACCTCG	660
QY	1983	tggagttctgtgaccagatgggctgcccattgagtgatgctcagtgaggggccctaaata	2042
Db	661	TGGAATTTCTGCGACCAAGTGGGCTGCCCGTGGACTTCAGCTCCGACGAGGCCCTCAATA	720
QY	2043	aaagcctgaactaagaactttggagatgacaagtagtctcttggcccggaggaggtgctca	2102
Db	721	AAAGCTGACCCAAACATTTGGAGATGACAAGTACTACTGGCCCGCCAGGAGGTGCTCA	780
QY	2103	ccaacatgactcccgcccatgacatggaactgtacttctgtctgtggatgctgcag	2162
Db	781	CCAACATGTCTCCCGGCCCATGCAGATGGCACGTGACTTCTGCTCGGGGCTGCTGCAGG	840
QY	2163	accagagcagttccggcattatgctctcaacgttccctctacacacactcaactctc	2222
Db	841	ACCCAGCGCAGTTCCGGCCTACTAGCGCTCAATGTGCCCTGTACACACACTTCACTGCG	900
QY	2223	ccatccgcgctttgtgacgtcatagtgaccgcctcctgctgctgcttgggctaca	2282
Db	901	CCATCCGCGCTTTGCCGACGTCTTGGTGCACCGCTCTCTGGCTGCCGCTTAGGCTATA	960
QY	2283	gtgaacagcgaagtgtgagccttgataccctacagaagaagctgacacctgaatgaacc	2342
Db	961	GGGAGCGACTAGACATGGCGCCGATACCTTCAGAAAACAGGCGACCACTGTAAAGACC	1020
QY	2343	gtcgcattggcttccaaactgtgacgagagctcagcatcgccctcttcttcgacgttctag	2402

Db	1021	GCCGATGGCTCCAAAGCGCTGCAGGAGCTCAGTACCAGTCTCTCTTTGCTGTCTGG	1080
QY	2403	taaagagagtgggccctcctggagtcocgaagccatggtgatggtgtcctgaaccaagctt	2462
Db	1081	TCAAGGAGAGTGGCCCTTGGAGTCAAGAAGCCATGTTGGGATCCTTGAAGCAAGCT	1140
QY	2463	tcgacgtgctgctgctgcttggggtcagaagcgcacatctactgcaatgcaactggccc	2522
Db	1141	TCGACGTGCTGCTGGCTACGGCTGCAGAGGGCATCTACTGCAAGCACTGGCC	1200
QY	2523	tgcgatctctcaagcttcagaaggtgggaagacccagagctcaactcttgtttgggaac	2582
Db	1201	TGCGGTCCCAACACTTCCAGAAAGTGGGCAAGAAGCCGAACTCACCTGGTGTGGAGC	1260
QY	2583	ctgatgacctgaagagagcacaacacagcaggtcatcaccatcttcagcctggtgagtg	2642
Db	1261	CTGAGGACATGGAGCAGGAGCCAGCACAGCAGTCAATCACCATCTTCAGCCTGTGGAGG	1320
QY	2643	tggtcctgcaggcagagccacagccctcaagtaagtaagtgctatcctgaagcagcagggcc	2702
Db	1321	TGTCCTCGAGGACAGTCCACAGCCCTCAAGTACAGCGCCATCCTGAAAGCGCCAGGCA	1380
QY	2703	t-----ggagaagcgctctgatgagagcctgaggaact	2735
Db	1381	CCCAGGGCCACCTGGGCCCTTGAGAAGGAGGAGGAGGTCTGACGGTGAGCCCGAGACT	1440
QY	2736	gaa-----tgctagccaagccagcctgtgctgcccctacccctgctggtt-----	2782
Db	1441	CAAGCACACAGCTGAGCTCCACACGCGCTTGCCTCCCGCTTGCCTGCTGCTGCCGCG	1500
QY	2783	---ttaggaataaggaccttttgacaccaaagggaatttttaatttggttttttaacaaactc	2839
Db	1501	ACACTGGCTTTAGGACCTTGTGACACGAGGGGGGTTTTTAATTTGGTTTTTAACAATC	1560
QY	2840	aggggttggtttttatttttttttcttttatttatttcttttgcagcagttttttaa	2899
Db	1561	AGGGGTTGTTTTTAT-----TTTTTAAATTTTTCAGCTCAACTTTTAA	1608
QY	2900	tgaactgggaaggttaggggtcagg--gcaggggatgctgagccctggcctgcttccct	2957
Db	1609	CAAACTGACGGGAGAGGGTGGGCTTGAAGAAAGCTGAGGCCCTGCTCAGCAGTACCC	1668
QY	2958	gagcagagagatccactcctcctgagcagcagcccgcttctaccagcagccact	3017
Db	1669	CAGCAGAGGAGCCCGCTCTCTGGAGGCTGCGCCCTTTTCTGG-CCCTACT	1727
QY	3018	gccttccctgccaggaaat-ggggggttttcagcaaatcagtcagtggaataaaa	3073
Db	1728	GCCCTCTCTGCCAGAAATGGGGGGTTTCAGCAACTCAGTGTACAGAAATAAA	1784

RESULT 4

AX015901	AX015901	Sequence 3 from Patent WO9950284.	1140 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX015901	Sequence 3 from Patent WO9950284.				
DEFINITION	AX015901					
ACCESSION	AX015901					
VERSION	AX015901.1	GI:10041643				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE	1 (bases 1 to 1140)					
AUTHORS	Aligner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.					
TITLE	Nucleic acid molecules which code proteins influencing bone development					
JOURNAL	PATENT: WO 9950284-A 3 07-OCT-1999;					
	ATGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP					
	ANDREAS (DE); WIRTH THOMAS (DE)					
FEATURES	Location/Qualifiers					
source	1..1140					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					

```
CDS
<2.784
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC07619.1"
/db_xref="GI:1004164"
/translation="IRAFPEQALLRRHPPTQRLMSDLVEFDQMGLPVDFFSAGAL
NNAIFYCSGLLDQPAQFRHYALNVLYFTSPIRFADVLVRLHLLAALGYRERLDM
APDITLQADHCNDRRMASKRVOELSTSLFVLVKSGLPSEANVMGILKQAFDVL
VLRYGOKRIYCNALRSHHFQVKRKPELTLLWEPEDMEQPEPAQOVITIFSLVEVY
LQAEStALK7SAILKRPQTQGHLPKEBESDGEPEDS7S"
BASE COUNT      231 a      355 c      334 g      220 t
ORIGIN
Query Match      18.2%; Score 565.6; DB 6; Length 1140;
Best Local Similarity 73.0%; Pred. No. 5.8e-136;
Matches 905; Conservative 0; Mismatches 189; Indels 146; Caps 8;
QY 1906 gatcttcgcacgttcctccctgagcagcgcctgctgcgcgcgcacatcccccacacagacgaa 1965
Dbb 1 GATCCACCGCGCTTCCCGAGCAGCGCCTGCTGCGCGCGACCCCGCCGCCAACAG 60
QY 1966 gatgctcaatgacctgggtggagttctgtgaccagatggggctgcacatggtatgctc 2025
Dbb 61 GATGCTCAGTGACCTGGTGGAAATTCGCGACACAGATGGGCTGCGCGTGGACTTCAGTCT 120
QY 2026 tgacggggcctaaataaaagcctgactaagacattggagatgacaagtactctctggc 2085
Dbb 121 CGCAGGAGGCCCTCAAT-----136
QY 2086 ccgggaaggagggtgctcaacaacatgtactcccgcccatgacatgagcagtgacctgtactctg 2145
Dbb 137 -----ATGGCACTGTACTTCTG 153
QY 2146 ctctgggatgctgcgagaccaggagcagttccggcattatgctctcaacgttccctcta 2205
Dbb 154 CTCGGGCTGCTGCGAGGACCCAGCGCAGTTCGCGCACTACGCGCTCAATGTGCCCTGTA 213
QY 2206 cacacactcaactccatccacccgccttctgctgacgtcactagtgacacgcctcctggc 2265
Dbb 214 CACACACTTCACTCGCCCATCCGCGCTTTGGCGACGTCCTGTGACCGCTCTCTGGC 273
QY 2266 tctgctctgggtacagtgaacacgacagatgtggagcctgataccctacagaaagcagc 2325
Dbb 274 TGCCGCGTTAGGCTATAGGAGCGACTAGACATGTGGCGCCGATACCTTGCGAGAACAGGC 333
QY 2326 tgacactgcaatgacctgcgatgcttccaaacgctgacgagctcagcagctcgcgctc 2385
Dbb 334 GGACCACTGTAACGACGCGCGCATGCGGTCCAAAGCGGTGCGAGGCTCAGTACCACTCT 393
QY 2386 ctcttcgcagttctagtaaaaggagagtggtgcccctcgagtcacgaagcattggtgaggg 2445
Dbb 394 CTCTCTTTGCTGTCTGTCGAAGGAGAGTGCGCCCTGGAGTCAGAGCCATGTTGATGG 453
QY 2446 tgcctgaaacaagcttgcagctgctggtgctgcgcttctggtgggtgcagaagcgcata 2505
Dbb 454 CATCTGGAAGCAAGCCTTCGACGCTGCTGTGCTGCGCTACGCGCTGCGAGAGCGCATCTA 513
QY 2506 ctgcaatgacactggccctgcgactctacagcttccagaaaggtggggaagacagcagct 2565
Dbb 514 CTCGAACGCACTGGCCCTCGGCTCCACCACTTCCAGAAAGTGGCGCAGAGCGCGAATCT 573
QY 2566 cactctgtttggagcctgatgacctgaagaggagcaacacagcaggtctacacat 2625
Dbb 574 CAGCCTGCTGAGGAGCCTGAGGACATGGAGAGGAGCGCACAGCAGTCTATCACCAT 633
QY 2626 ctccagcctgggtgagtggttctctgcagcagagggccacagccctcaagtacagtgctat 2685
Dbb 634 CTTCAGCCTTGGTGGAGTGTGCTTCGCGAGCAGAGTCCACAGCCCTCAAGTACAGCCCAT 693
QY 2686 cctgaagcagacaggc-----ctggaagagcgcctcta 2718
Dbb 694 CTTGAAGCGGCCAGCGACCCAGGGCCACCTGGGCGCTGAGAAGAGGAGGAGGAGTCTGA 753
```

```
QY 2719 tgagagcctgagactgaa-----tgctagcccaagccagggcctgtgctgcctacacc 2773
Dbb 754 CGGTGAGCCCGAGGACTCAAGCACCAGCTGCCAGCCGCGCTGCCCGCGCTGCC 813
QY 2774 tgcgtgctt-----ttaggaatagacacctttgacacacaaagggtattttaat 2822
Dbb 814 CGCCTGCCTGTCGCCGCCACACTGGCTTTAGGACCTGTGTGACACGGAGGGGTTTAAAT 873
QY 2823 ttggttttaaacactcagggttggtttttttttttttttttttttttttttttttttttt 2882
Dbb 874 TTGGTTTTTAACAACCTCAGGGGTTTGTTTTATTATTATTA-----ATTTTT 921
QY 2883 gcagctcagtttttaaatgaactggaaggttaggggtcagg--gcaggggagtgctgaggc 2940
Dbb 922 GCAGCTCAACTTTTAAACAACATGCAGGGAGAGGGTGGGCTTGAAGAGAGCTGAGGC 981
QY 2941 ctggcctgtgcttcctgcagcagagagatccagtcctcctctctctctctctctctctct 3000
Dbb 982 CTGTCAGCAGTGAACCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1041
QY 3001 ctaccagcgaccactgccttccctccctccctccctccctccctccctccctccctccct 3059
Dbb 1042 TTTTCTGG-GCCTTACTGCTCTCTGCGCAGGAAATGGGGGGGTTTCAGCAACTCAGT 1100
QY 3060 gtcagtgaataaaatcaatgtgaaaaaataaaaaa 3099
Dbb 1101 GTACAGATAAATAATCAAGTGTGGAGTGCCATAAAAAA 1140
RESULT 5
AB066536 2189 bp mRNA linear PRI 26-JUL-2001
LOCUS Macaca fascicularis brain cDNA clone:QmoA-12105, full insert
DEFINITION
AB066536
sequence.
ACCESSION AB066536
VERSION AB066536.1 GI:15021873
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Macaca fascicularis adult male medulla oblongata cDNA to mRNA,
clone.lib:macaque brain cDNA library QmoA clone:QmoA-12105.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.
1 (bases 1 to 2189)
Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished
2 (bases 1 to 2189)
Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
Submitted (23-JUL-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/,
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DralII (CACTGTGTG)
R. Site2: DralII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGTGGCGCTTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5Kb.The SfiI-digested PCR product was cloned
into distinct DralII sites of pME18S-FL3. XhoI sites just outside
the DralII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al.(University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
( 5' end primer [CTTCTGCTCTAAAGCTGCG] ;
3' end primer [CGACTGCGAGTCGAGACA] ).
```





Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M. Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence Unpublished  
 2 (bases 1 to 82454)  
 Lin, X. and Kaul, S.  
 Direct Submission  
 Submitted (21-OCT-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org  
 3 (bases 1 to 82454)  
 Town, C.D. and Kaul, S.  
 Direct Submission  
 Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
 On Jan 19, 2001 this sequence version replaced gi:12280880.  
 Address all correspondence to: at@tigr.org  
 BAC clone T32E8 is from Arabidopsis thaliana chromosome 1  
 The orientation of the sequence is from SP6 to T7 end of the BAC clone.  
 Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://genemark.biology.satech.edu/GeneMark/>), GlimmerA (a variant of Glimmer, see Mihaela Pertea, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).  
 Location/Qualifiers  
 1..82454  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="T32E8"  
 1..189  
 /note="overlap with BAC clone T5M16  
 [AC010704:103165..103353]"  
 complement(join(<1..69,154..2280,2381..2449,2524..2577,2676..3055,3215..3320))  
 /gene="T32E8.1"  
 complement(join(<1..69,154..2280,2381..2449,2524..2577,2676..3055,3215..3320))  
 /gene="T32E8.1"  
 /codon\_start=1  
 /product="putative 3'-5' exoribonuclease, 3' partial;  
 3320-1"  
 /protein\_id="AG51632.1"  
 /db\_xref="GI:12323300"  
 /translation="MKSASSEQSVERTENGKKRNRFQKQNRKSSQSVPIEDAHVE  
 ESDGRDSSRKADSTSSKQRPNTDELEAMRASNVAFNSMPPMRAESGYPRRSAS  
 PLISSPEVSKLLSKSPDPRACEQSPGMNGLFQIEGSSQRIKFSHSLSDAVTEA  
 LKGEAFKALFRVNAHNRNEAYCKIDQPTDILINGVQCSRAVEGTVVILKIDPLSL  
 WPMKGFVTEAAKPECTNSPPPEKDDKKAROKINDVVEGEDGFSGKNSVIGKCAK  
 NGVTPSSPSLDSLCFSFCQKGNCSAVDKCLILSSPEKRPQGVAVVEKSLYRD  
 SYGLLDVNLAEAGVAAQIVDMGEGSPFVAQITHLFGSGELEFPQINAILYQNSVCD  
 ARLENDVNLAEAGVAAQIVDMGEGSPFVAQITHLFGSGELEFPQINAILYQNSVCD  
 SDFSGSTSLPRVPEVPEEVRORRDLCLVLTIDFTDLDLALSQSLPGGF  
 RYFGVIAVDSYFVLPTALDTEARFSTSVLMQRKISMLPPLLSNVGSLSPGADR  
 LAFSILMDLNRGQDVIDWIGRTIIRSCCKLSYDHAQDIDGKSDVAENGPALHGSF  
 KWCDVTRSVKQLSEISTTLRQKRFNGALQLENSKPVFLFDEHGVDPYDFVTCRSKGSN

FLVEFMLANNTAEVISOAYRASSILRRHPNTRKLKEFEGFCKSKHGMDLDISSS  
 GQDSELEKTNLKDVSFVDVILNNYAKPQLASVCTGNLKDVSVAENGHYALAVP  
 LYHFTSPURRYPDIVHRAALAAELIYSKQKOTALDEGSCFTGIIHFNKDAES  
 IEGKEALNSVGLKMGVSPSTELSDVAAYCNERKLAARKVRDACDKLYTFWVLAQKEIF  
 PCEARVNLGSRFMTVYISKL"  
 complement(1..3320)  
 /gene="T32E8.1"  
 /note="similar to ribonuclease II RNB family protein  
 GB:CA816367 [Schizosaccharomyces pombe]; Pfam HMM hit:  
 vacB and RNase II family 3'-5' exoribonucleases"  
 complement(join(<8589..8783,9000..9465,9555..9652,  
 9753..9936,10020..10132,10236..10421,10504..10716))  
 /gene="T32E8.2"  
 complement(8589..10716)  
 /gene="T32E8.2"  
 /note="similar to AUX1 (regulator of root gravitropism,  
 putative permease) GB:CAA67308 [Arabidopsis thaliana]"  
 complement(join(8589..8783,9000..9465,9555..9652,  
 9753..9936,10020..10132,10236..10421,10504..10674))  
 /gene="T32E8.2"  
 /codon\_start=1  
 /product="putative AUX1-like permease; 10674-8589"  
 /protein\_id="AAG51630.1"  
 /db\_xref="GI:12323298"  
 /translation="NAAEKIEVVAGNVLEMEEREENISGNKSKSTKLSNFWHGG  
 SYDNWESCASNOVAQVLLTLPYSFSGLMGGLFQLEFGLMGSWATYLLISLVYVY  
 RTKREKEDFRNHVQLQFEVLDGLGKHWNGLJFNGTFLFGSVTLQIACASNIY  
 YINDKLDKRTYIIFGACATVTFIPSFHRYRWSFGLAMTYTSWLTATSLHGO  
 AEDVSGGPTVWLTGATNLTIFGGHAVTVIEHMKRQKFKALYLLATYIVLT  
 LTPSASVYWAFGDKLLTHSNALSLLPKTFGRDTAVILMLHIHQITFGFATPLYFV  
 WEKLIQVHETKSMFKRAMRLPVVPIWFLAIIFFPFGINSAGSLVSTVYIIPA  
 LAHMLFAPAPSRANAVPRPVVGMWGYTCINIEVVVWVFGVFGGWSMVNFV  
 RQIDTFGLTKCYCPHPKP"  
 join(<12104..12179,12505..12628,12704..13574)  
 /gene="T32E8.3"  
 12104..13574  
 /gene="T32E8.3"  
 /note="similar to thaumatin-like protein GB:AAB71214  
 [Arabidopsis thaliana]; Pfam HMM hit: Thaumatin family"  
 join(12104..12179,12505..12628,12704..13574)  
 /gene="T32E8.3"  
 /codon\_start=1  
 /product="thaumatin-like protein; 12104-13574"  
 /protein\_id="AAG51631.1"  
 /db\_xref="GI:12323299"  
 /translation="MDRTIEIEFDINGPYSLRPNLSVSLTSSENITFDPKIES  
 SKDANSFRFVKDLVTLSHIIRVMTITRTPVNSTGKWSERARFTIISCCQTIW  
 PAITPGENFGGFGELKPGQSIIVHAPVGSGRITGRTGCKFDSTGTCTGTSCGST  
 LKCSAGKPPASIAEFTLAALDFYDVLVDGFLNPSVTPMNGKGNCSVAGCVADLRP  
 HCPQLAVKNGKVIISCSACDVFDREYCCRGVGNPVQCPYYSKIFKQACPTAY  
 SYAYDDPTSIMTASDVIYSPCSSRFLVFTLTSHDYFNVDNSQVSFILHDHFYIY  
 MIYTHPLLISYAKVKVIMOK"  
 join(14054..14174,14251..14348,14779..14811,15170..15303)  
 /gene="T32E8.4"  
 14054..15303  
 /gene="T32E8.4"  
 /note="similar to hypothetical protein GB:P34661  
 [Caenorhabditis elegans]"  
 join(14107..14174,14251..14348,14779..14811,15170..15252)  
 /gene="T32E8.4"  
 /codon\_start=1  
 /product="unknown protein; 14107-15252"  
 /protein\_id="AAG51633.1"  
 /db\_xref="GI:12323301"  
 /translation="MATGKGVSKVLTSDPKLPKVFSPYEGAPFATVLKFAAEFR  
 VPPTSAITINDGIGINPQOSAGNVFLKHSSELRLIPDRVGAFFVMDP"  
 join(<18583..18916,19056..19948,20068..20268,20382..20543,  
 20662..21389)  
 /gene="T32E8.5"  
 18583..21386  
 /gene="T32E8.5"  
 /note="Pfam HMM hit: Eukaryotic protein kinase domain,  
 C-terminal region is identical to putative protein kinase

CDS  
GB:CAA72680 [Arabidopsis thaliana]  
Join(18583..18916,19056..19948,20068..20268,20382..20543,  
20662..21234)  
/gene="T32E8.5"  
/codon\_start=1  
/product="putative protein kinase; 18583-21234"  
/protein\_id="AAG51619.1"  
/db\_xref="GI:12323287"  
/translation="MOTTISGPRRSVAPQROASRNLRLVTAEGORSQDVVLTQSILAA  
NFLQDTNNLATISVAGESASLTQPTVSEHNPDRDPGKSAVTSLESNLDVQRK  
SOSLIGTSQDMDENDATNOAESHLDTACGSKHQLPSYDSEVLSKSEYKDSLSLAKIQ  
GOLGEFNEINOPRTRCSAVGSNAITLIIHSSAPMLNATVHSKSYVEADSNANPH  
AVQSGNLPSCDSKSVSNLHPNDKASEPASTNDPEVRKEDTSKQOQITGL  
EAPVSSIIYSDQANARPEELHTSVSSQPKDKHKEVASSGPSAPRKRYDPDL  
FKVNGKLIORLKGSGSSVHKVISEDCTIYALKIKLGRDVTATYGFQCEIGY  
LKKKNTNIQILDYEDTKTLLQVLNGTMSNKGVRVEDGFIVYVLEIGEIDLAH  
MLSKWREIEGSDRTIDENLRFYQQIOLQAVNTIHEERIVHSDLPANFLVLRGFLK  
LIDFGIAKANSOTNTIQRDSOVFTLSYMSPEAFMCESDENGNTIKCCRPSDWSLG  
CITLQMVYGRTPADYKTFWAFKVTIDPNHEITVYNOLSNPWLIDLMKCLANDBNQR  
WRIPQLQHPPLAPPPIPEPOVKTIKFLSLIAESCGSDDDKANSMSISLEQLLSNPAP  
LPRNDVLSRQNOQLLSRVSELCIQDLQGLGEN"  
Join(<21802..22203,22526..>22921)  
/gene="T32E8.6"  
/note="22921  
21802..22921  
/gene="T32E8.6"  
/note="similar to putative SWH1 protein GB:AAC20736  
[Arabidopsis thaliana]; Pfam HMM hit: PH (pleckstrin  
homology) domain"  
Join(21802..22203,22526..22921)  
/gene="T32E8.6"  
/codon\_start=1  
/product="hypothetical protein; 21802-22921"  
/protein\_id="AAG51621.1"  
/db\_xref="GI:12323289"  
/translation="MSSPLAAKSNLTTRASLNIPHSRTHNHDRLSLSAGPRTN  
EDRPSAGNVAGILYKWNVGGKRRFVLDQGLVLSYRIHPDPKISLSVEMDRSK  
LIGESLRFICRSGKRGDVHSPKPLGOIHLKVSIGOSDKGRTPTVGTGKSLHR  
AATSEDRSATEALKAVKETFPRMSNEELMNTSVNSVDKLRQRLMEVEDEITLK  
DCEDIMKNFIALHDRVMSLKRYQTHLIDLSKVNSPRPSQSF.  
Join(217895..279331,27952..29091,29174..29338,29453..29618,  
29775..29885,29967..30160,30255..30385,30469..30562,  
Query Match 8.08; Score 247.4; DB 8; Length 82454;  
Best Local Similarity 51.98; Pred. No. 8.8e-53;  
Matches 657; Conservative 0; Mismatches 596; Indels 12; Gaps 4;  
Qy 1015 gctgttcacatcgcacatagactgaagaggactgtatatttgcctcggggcaact 1074  
Db 1675 GCTGTAGCTGCCCAATTTAGACTGGGCGAAGGAGTCCATTTCTGTAGCCCAAT 1616  
Qy 1075 ggcctagagatcttgaggcgggtggaatcgagctgaacagaaggggaactgacaga 1134  
Db 1615 TACACATCTATTGGCCGAGCGACGTAAATTGGAGCTCAGATCAATGCGATATTATCA 1556  
Qy 1135 atatgtgtggagactctctatctcttcacaaagtcttgaagtctcctcctcaagcct 1194  
Db 1555 GAACTCGGTTTCGGATTCTGACTTCTCCAGATCGCTTACTAGTCTTCCAC---GTGT 1499  
Qy 1195 gccctggacaatccacatgatagggtgggcaagagagagacacctaaggaagaaactgtat 1254  
Db 1498 GCCTGGGAAGTACCAGAAGAAGAGTTCAACGTAGAAAAGATCTGAGAGACCTGTGTGT 1439  
Qy 1255 cttcacattgatccatcactctcgcacattgaatgatccctcctgcaggcggt.1314  
Db 1438 ATTACCATTCACCTTCGACAGCTACAGATCTCGATGATGCTCTATCAGTGCMAAGTTT 1379  
Qy 1315 cactgatggcactctgaagtggcgctccacatcgcggatggagttacttttctctga 1374  
Db 1378 ACCCGCGGTTTTTTTATAGTCGCGTGTTCACATTGCTGATGCTCTCTCTCTTACC 1319  
Qy 1375 gggatcctctttgataaagttagctgtgagagagcacaagtgtactgttcacagaa 1434  
Db 1318 AGAGACTGCTCTGCACACAGAAGTCGGTTTAGTCAACCACTGTTTACTTTGATGCAGAG 1259

Qy 1435 ggtgttccccatgcttccaggcttctgtgtgaggaactctgcagccctcaaccccatgac 1494  
Db 1258 AAAAAATATCAATGTTGCCCTCCATTACTGTCTAGAGAAACGCTAGGTTCTAGTCCAGGAGC 1199  
Qy 1495 tgacaagctgacctctctgtgtgagcctgacccctgaagcgaagcgaagcctcttgaaga 1554  
Db 1198 TGATAGACTTCCTCTTTTCTATCTTTGGGATTTAAATAGAGAGGGATGTGATAGATCG 1139  
Qy 1555 gtggtttggccgactatcctcgttctgtgcacaaactgagctaacgacatgccccagag 1614  
Db 1138 CTGCGATTTGGCGTACTACTACATACGTCGTCGTCGAAGCTTTCATATGACCATGCTCAAGA 1079  
Qy 1615 catgatcgaataatcccaactgagaagatccctcctgaggaagccttccccaatctccaga 1674  
Db 1078 CATATATAGATGGCAAAAAGTG--ATGTTCCAGAAAATGGCTGGCCTGCGTTCATGGATC 1022  
Qy 1675 gcaagcgtcgaggaggtgcaccaggcagctcctgaactgcacagcattgcacaaact 1734  
Db 1021 TTTCAATGSGTGATGTAACTCGATCTGTCAAAACAGCTTAGTGAGATTTCCACCACCTTT 962  
Qy 1735 ccgcgcgcagcgcttctgtagatggcgactcctcgttttagatcagctgaagcttgcctttac 1794  
Db 961 AAGGCAAAAAGATTTAGAAATGGGCTTTGCAGCTGGAGAAATTCGAAGCTGTTTT--- 903  
Qy 1795 tctggaccatgagactgagctgcctcgaaggtatgcacatctatgataccgagacagcaa 1854  
Db 904 TCTATTGATGACATGGAGTTCATATGATTTTGTGACGTGTTCTAGAAAAGGGTTCGAA 845  
Qy 1855 caagctgtagaggagttcatgctcctcgtggccaaactggcgggtggcccaagaatctccg 1914  
Db 844 TTTTCTGTTGAGGAGTTTATGCTCTTAGCAAAATATCACAGCAGCTGAGGTTATTTCTCA 785  
Qy 1915 cacttccctgagcggcctcctcgtcgcggcatcccccacacacagaagatgctcag 1974  
Db 784 ACCTTACCGTGTAGCTCATTTGCTCCGAGGACATCCAGAACAAATACCTGCGCAAGCTCAA 725  
Qy 1975 tgacctggtggagttctgtgacagatgggctgcccatggtatgctcagctgtcagctggagggc 2034  
Db 724 AGAATTTGAAGGGTTTTTGTGTTAAAGATGATGATTTGGACATTTCTCTCTGCGCA 665  
Qy 2035 ctaaaataaagcctgactaaagacatttggagatgacaaactcctcgtgccggaagga 2094  
Db 664 ACTTCAGATTCATTGGAGAAAATCAGTGGAAACCTCAAGAGTACGACTAGTTTTGTGGA 605  
Qy 2095 ggtgctcaccacacatgtactcccggcccatgcagatggcactgtctctcctctggat 2154  
Db 604 TATTCTCAACAATTTATGCTATAAAACCTATGCAAACTAGCATCTTACTCTGCACTGGAA 545  
Qy 2155 gctcgagcaccaggagc---adgtccggcattatgctcctcaacgttcccctctacacaa 2211  
Db 544 TTTGAAGGACAGTGTGCTGAGTGGGGACACTATGCACTAGCTAGTGTCTCTCTATCTCA 485  
Qy 2212 cttcacctctccatccgcgcgttctgtcagctcatagtgcaaccgctcctcctcgtcgtgc 2271  
Db 484 CTTACATCTCCTCTTCGACGCTACCCAGACATAGTCTGTTACCGGTGCTTGTAGTGTGTCG 425  
Qy 2272 tctgg 2276  
Db 424 TCTTG 420

RESULT 10  
AX015902  
LOCUS AX015902 4999 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 5 from Patent WO9950284.  
ACCESSION AX015902  
VERSION AX015902.1 GI:10041645  
KEYWORDS house mouse.  
SOURCE Mus musculus.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```
REFERENCE 1 (bases 1 to 49999)
AUTHORS Aligner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 5 07-OCT-1999;
ALIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
source
location/Qualifiers
1..49999
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 13210 a 11814 c 10825 g 14150 t
ORIGIN
Query Match 7.6%; Score 236.8; DB 6; Length 49999;
Best Local Similarity 99.2%; Pred. No. 4.6e-50;
Matches 238; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1542 agatccttgagagtggtttggccgcactatccgttcttgacccaactgagctacg 1601
DB 8518 AGATCCTTGAAGAGTGGTTGGCCGCACATATCATCCGTTCTTCACCAAACTGAGCTACG 8577
QY 1602 accatgccagagcatgatcgaaaataccaactgagaagatccctgaggagagcttcccc 1661
DB 8578 ACCATGCCAGAGCATGATCGAAAATCCAATCGAAGATCCCTGAGGAAGAGCTTCCCC 8637
QY 1662 caattctccagagcacagcgtcgaggagggtgcaccaggcagtcctgaaacctgcacagca 1721
DB 8638 CAATTTCTCAGAGCACAGCGTCGAGGAGGTGCACCAGGCAGTCCCTGAACCTGCACAGCA 8697
QY 1722 ttgcaaaagcaactcccccgcagcgtttgtagatgcgcactcccttttagatcagctga 1781
DB 8698 TTGCAAAAGCAACTCCCGCCAGCGCTTTGTAGATGGCGCACTCCCTTTAGATCAGGTCA 8757
RESULT 11
LOCUS AX015908 49999 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 11 from Patent WO9950284.
ACCESSION AX015908
VERSION AX015908.1 GI:10041651
KEYWORDS house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 49999)
AUTHORS Aligner,T., Hess,J., Rosenthal,A., Rump,A. and Wirth,T.
TITLE Nucleic acid molecules which code proteins influencing bone
development
JOURNAL Patent: WO 9950284-A 11 07-OCT-1999;
ALIGNER THOMAS (DE); HESS JOCHEN (DE); ROSENTHAL ANDRE (DE); RUMP
ANDREAS (DE); WIRTH THOMAS (DE)
FEATURES
source
location/Qualifiers
1..49999
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 13135 a 11787 c 10868 g 14209 t
ORIGIN
Query Match 7.6%; Score 236.8; DB 6; Length 49999;
Best Local Similarity 99.2%; Pred. No. 4.6e-50;
Matches 238; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1542 agatccttgagagtggtttggccgcactatccgttcttgacccaactgagctacg 1601
DB 10503 AGATCCTTGAAGAGTGGTTGGCCGCACATATCATCCGTTCTTCACCAAACTGAGCTACG 10562
QY 1602 accatgccagagcatgatcgaaaataccaactgagaagatccctgaggagagcttcccc 1661
DB 10563 ACCATGCCAGAGCATGATCGAAAATCCAATCGAAGATCCCTTGAGGAGAGCTTCCCC 10622
```

```
QY 1662 caattctccagagcacagcgtcgaggagggtgcaccaggcagtcctgaaacctgcacagca 1721
DB 10623 CAATTCTCCAGAGCACAGCGTGGAGGAGGTGCACAGGCAGTCCCTGAACCTGCACAGCA 10682
QY 1722 ttgcaaaagcaactcccccgcagcgtttgtagatgcgcactcccttttagatcagctga 1781
DB 10683 TTGCAAAAGCAACTCCCGCCAGCGCTTTGTAGATGGCGCACTCCGTTTATAGATCAGGTCA 10742
RESULT 12
LOCUS AY058669 3597 bp mRNA linear INV 17-OCT-2001
DEFINITION Drosophila melanogaster LD37985 full length cDNA.
ACCESSION AY058669
VERSION AY058669.1 GI:16198178
KEYWORDS FLI_CDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 3597)
REFERENCE Stapleton,M., Brokstein,P., Hong,L., Agbavani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Friese,E., George,R.,
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
Nunoo,J., Pacieb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
Direct Submission
Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Lawrence Berkeley National Laboratory
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones, that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1..3597
/organism="Drosophila melanogaster"
/strain="Y; cn bw sp"
/db_xref="taxon:7227"
/map="61A6-61A6"
1..3597
/gene="CG16940"
/note="alignment with genomic scaffold AE003467"
/db_xref="FLYBASE:FBgn0035111"
109..3207
/gene="CG16940"
/note="Longest ORF"
/codon_start=1
/db_xref="FLYBASE:FBgn0035111"
/product="LD37985p"
/protein_id="AAL13898.1"
/db_xref="GI:16198179"
/translation="MSETSSVNVDDRCRIEGLRORSORLARLQMQRMESNAS
SKDLSANNAVLPQLPWRWAEPRPNQSPQSAKNSQAIPIYGVPTPAPVAMOLP
YHVPASTQATMOTPCILPILSHLVVSSPSHRLIAKKKRLRNLVEALQDMVKHLS
HLSPDYNVACQLIRKNEILSOLEWSQTRSEAHFPQVGYQGRHHDSEVATNSGSHV
```



## AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Blinag,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinch,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulsegad,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,  
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
Oguth,M., Okunoye,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,  
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs/R.

## TITLE

JOURNAL  
REFERENCE  
2 (bases 1 to 153701)

Worley,K.C.

Direct Submission

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17064556.

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GIHP  
Center clone name: CH230-145D14  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList  
Consensus quality: 123669 bases at least Q40  
Consensus quality: 129396 bases at least Q30  
Consensus quality: 135066 bases at least Q20  
Estimated insert size: 124661; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 61 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 8487: contig of 8487 bp in length  
8488 8587: gap of unknown length  
8588 12859: contig of 4272 bp in length  
12859 12959: gap of unknown length  
12959 17299: contig of 4770 bp in length  
17299 17829: gap of unknown length  
17829 23516: contig of 5687 bp in length  
23516 23617: gap of unknown length  
23617 27189: contig of 3573 bp in length  
27189 27289: gap of unknown length  
27289 31122: contig of 3833 bp in length  
31122 31223: gap of unknown length  
31223 35679: contig of 4457 bp in length  
35679 35775: gap of unknown length  
35775 35780: contig of 3290 bp in length  
35780 39069: gap of unknown length  
39069 39169: contig of 4254 bp in length  
39169 43423: gap of unknown length  
43423 47921: contig of 4398 bp in length  
47921 48021: gap of unknown length  
48021 51363: contig of 3342 bp in length  
51363 51463: gap of unknown length  
51463 54533: contig of 3076 bp in length  
54533 54640: gap of unknown length  
54640 57365: contig of 2725 bp in length  
57365 57464: gap of unknown length  
57464 61911: contig of 4447 bp in length  
61911 62012: gap of unknown length  
62012 65486: contig of 4475 bp in length  
65486 65877: gap of unknown length  
65877 69977: contig of 3391 bp in length  
69977 70077: gap of unknown length  
70077 73099: contig of 3022 bp in length  
73099 73100: gap of unknown length  
73100 76223: contig of 3024 bp in length  
76223 76323: gap of unknown length  
76323 78364: contig of 2041 bp in length  
78364 78465: gap of unknown length  
78465 80633: contig of 2168 bp in length  
80633 80733: gap of unknown length  
80733 83076: contig of 2344 bp in length  
83076 83176: gap of unknown length  
83176 86186: contig of 3010 bp in length  
86186 86286: gap of unknown length  
86287 89099: contig of 2813 bp in length  
89099 89199: gap of unknown length  
89199 91514: contig of 2315 bp in length  
91514 91614: gap of unknown length  
91614 93490: contig of 1876 bp in length  
93490 93590: gap of unknown length  
93590 95001: contig of 1411 bp in length  
95001 95101: gap of unknown length  
95101 97069: contig of 1968 bp in length  
97069 97169: gap of unknown length  
97169 99222: contig of 2053 bp in length  
99222 99322: gap of unknown length  
99322 102096: contig of 2774 bp in length  
102096 102196: gap of unknown length  
102196 103303: contig of 1107 bp in length  
103303 103403: gap of unknown length  
103403 103586: contig of 2183 bp in length  
103586 105587: gap of unknown length  
105587 107855: contig of 2169 bp in length  
107855 107955: gap of unknown length  
107955 110106: contig of 2151 bp in length  
110106 110206: gap of unknown length  
110206 112025: contig of 1819 bp in length  
112025 112126: gap of unknown length  
112126 113623: contig of 1498 bp in length  
113623 113723: gap of unknown length

```
* 113724 115968: contig of 2245 bp in length
* 115969 116068: gap of unknown length
* 116069 118090: contig of 2022 bp in length
* 118091 118190: gap of unknown length
* 118191 119318: contig of 1128 bp in length
* 119319 119418: gap of unknown length
* 119419 120531: contig of 1113 bp in length
* 120532 120631: gap of unknown length
* 120632 122162: contig of 1531 bp in length
* 122163 122262: gap of unknown length
* 122263 124304: contig of 2042 bp in length
* 124305 124404: gap of unknown length
* 124405 126716: contig of 2312 bp in length
* 126717 126816: gap of unknown length
* 126817 127986: contig of 1170 bp in length
* 127987 128086: gap of unknown length
* 128087 129847: contig of 1761 bp in length
* 129848 129947: gap of unknown length
* 129948 131590: contig of 1643 bp in length
* 131591 131690: gap of unknown length
* 131691 133025: contig of 1335 bp in length
* 133026 133125: gap of unknown length
* 133126 134625: contig of 1500 bp in length
* 134626 134725: gap of unknown length
* 134726 135806: contig of 1081 bp in length
* 135807 135906: gap of unknown length
* 135907 137024: contig of 1118 bp in length
* 137025 137124: gap of unknown length
* 137125 138379: contig of 1255 bp in length
* 138380 138479: gap of unknown length
* 138480 139623: contig of 1144 bp in length
* 139624 139723: gap of unknown length
* 139724 140842: contig of 1119 bp in length
* 140843 140942: gap of unknown length
* 140943 142718: contig of 1776 bp in length
* 142719 142818: gap of unknown length
* 142819 143980: contig of 1162 bp in length
* 143981 144080: gap of unknown length
* 144081 145323: contig of 1243 bp in length
* 145324 145423: gap of unknown length
* 145424 146688: contig of 1265 bp in length
* 146689 146788: gap of unknown length

Query Match
Best Local Similarity 7.29; Score 222; DB 2; Length 153701;
Matches 228; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1542 agatccttgaagagtgttggccgcactatcatccgttcttgcacaaactgagctacg 1601
Db 94361 AGATCCTTGAGGAGTGGTTGGCGCAGTATCATCGTTCCTGACCAAAATGAGCTACG 94302

Qy 1602 accatgccagacgatgcgaaatccaaactgagaagatccctgaggaaagcttccc 1661
Db 94301 ACCATGCCAGACGATGATCGAAATCCGACTGAGAAGATCCCAAGGAAGAGTTCCTCC 94242

Qy 1662 caattctccagacgacgctcgaggaggtgcacaggcgadctctgacactgcacagca 1721
Db 94241 CTATTTCTCAGAGCACAGCATCGAGGAGTGCACCGAGCAGTCTCGAACCTGCACAGCA 94182

Qy 1722 ttgcaagcaactccgcgcagcgtttgtatagatggcactccgtttagatcagct 1779
Db 94181 TTGCAAGCACTCCGCCGCCAGCGCTTGTAGATGGTGCATCTCGTTTAGATCAGGT 94124

RESULT 14
AC024062/c Homo sapiens chromosome 2 clone RP11-95H14, WORKING DRAFT SEQUENCE,
LOCUS AC024062 162842 bp DNA linear HTC 10-MAR-2001
DEFINITION 7 unordered pieces.
ACCESSION AC024062
VERSION AC024062.27 GI:13270576
KEYWORDS HTG; HTGS_PHA5E1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
```

```
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162842)
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federspiel,N., Glukhov,S., Hansen,N., Hyman,Z.S., Hyman,R.,
Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelm,J.,
Yu,S. and Davis,R.W.
Unpublished
2 (bases 1 to 162842)
Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N.,
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelm,J.,
Yu,S. and Davis,R.W.
Direct Submission
Submitted (22-FEB-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Mar 10, 2001 this sequence version replaced gi:13236644.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDSTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: 842
Center clone name: RP11-95H14
----- Summary Statistics
Sequencing Vector: M13mp18; X02513; 99% of reads
Sequencing Vector: plasmid; plasmid_accession: 1% of reads
Chemistry: Dye-terminator; 2% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 159366 bases at least Q40
Consensus quality: 159909 bases at least Q30
Consensus quality: 160434 bases at least Q20
Insert size: 166936; agarose-fp
Insert size: 162242; sum-of-contigs
Quality coverage: 14.6x in Q20 bases; agarose-fp
Quality coverage: 15.0x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1779: contig of 1779 bp in length
* 1780 1879: gap of unknown length
* 1880 11374: contig of 9495 bp in length
* 11375 11474: gap of unknown length
* 11475 38208: contig of 26734 bp in length
* 38209 38308: gap of unknown length
* 38309 65551: contig of 27243 bp in length
* 65552 93861: gap of unknown length
* 93862 93981: contig of 28230 bp in length
* 93982 127838: contig of 33857 bp in length
* 127839 127939: gap of unknown length
* 127939 162842: contig of 34904 bp in length.
FEATURES
Location/Qualifiers
1..162842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-95H14"
/clone_lib="RPCI human BAC library 11"
1..1779
/misc_feature
/note="assembly_name:Contig59"
1880..11374
```

misc\_feature /note="assembly\_name:Contig61"  
11475..38208  
misc\_feature /note="assembly\_name:Contig62"  
38309..65551  
misc\_feature /note="assembly\_name:Contig63"  
clone\_end:SP6"  
65652..93881  
misc\_feature /note="assembly\_name:Contig64"  
93982..127838  
misc\_feature /note="assembly\_name:Contig65"  
clone\_end:17"  
127939..162842  
misc\_feature /note="assembly\_name:Contig66"  
127939..162842  
BASE COUNT 45634 a 33939 c 33845 g 48820 t 604 others  
ORIGIN

Query Match 6.9%; Score 215.4; DB 2; Length 162842;  
Best Local Similarity 89.9%; Pred. No. 2.1e-44;  
Matches 231; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 819 aggtgtttacatcttgagaaaagcattctcgagcagcaactggcattctcttccttcagaccct 878  
|||||  
Db 107889 AGGTGTTTACATCTTGGAGAAAAACATTTCTCGAGCAGCAACGGCTTCTCTCAAACTCT 107830  
|||||

QY 879 tggctgataagacagtcagctgtttaagaaaatacgcctctgttttctcttcagaccacc 938  
|||||  
Db 107829 TGGCTGTATAGACAGCGAACTGTTTAGGAAATACGCCCTGTCTCTCCCTCAGACCACC 107770  
|||||

QY 939 gagtacctagaattacgtacctctcaaggactgtcccgaggactcaccgaccacct 998  
|||||  
Db 107769 GAGTGCCTAGAAATTTATGTGCTCTCAAGGACTGTGCCAGGACTTTGTGGCACGGCCTA 107710  
|||||

QY 999 aagacttgcacacagctgttcattcgcgcacatagactgggaaggaggactgtaatt 1058  
|||||  
Db 107709 AAGATTATGCCAACACACTGTTTCATCTGCCGCATTGTGGACTGGAAGGAGGACTGCAATT 107650  
|||||

QY 1059 ttgcctctggggcaactg 1075  
|||||  
Db 107649 TTGCCCTGGGGTAGGTG 107633  
|||||

RESULT 15  
AF312225  
LOCUS AF312225 345102 bp DNA linear PRI 17-JAN-2002  
DEFINITION Homo sapiens chromosome 2 map 2q37.1, complete sequence.  
ACCESSION AF312225  
VERSION AF312225.1 GI:18182287  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 345102)  
AUTHORS Rump, A., Hess, J., Wen, G., Wirth, T. and Rosenthal, A.  
TITLE Human genomic sequence from 2q37.1  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 345102)  
AUTHORS Rump, A., Hess, J., Wen, G., Menzel, U., Wirth, T. and Rosenthal, A.  
TITLE Direct Submission  
JOURNAL Submitted (11-Oct-2000) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
FEATURES  
source  
1. 345102  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2q37.1"

BASE COUNT 92011 a 70334 c 74209 g 108499 t 49 others  
ORIGIN

Query Match 6.9%; Score 215.4; DB 9; Length 345102;  
Best Local Similarity 89.9%; Pred. No. 2.1e-44;  
Matches 231; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Best Local Similarity 89.9%; Pred. No. 2.4e-44;  
Matches 231; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 819 aggtgtttacatcttgagaaaagcattctcgagcagcaactggcattctcttccttcagaccct 878  
|||||  
Db 153908 AGGTGTTTACATCTTGGAGAAAAACATTTCTCGAGCAGCAACCGGCTTCTCTCAAACTCT 153967  
|||||

QY 879 tggctgataagacagtcagctgtttaagaaaatacgcctctgttttctcttcagaccacc 938  
|||||  
Db 153968 TGGCTGTATAGAACACAGCGAACTGTTTAGGAAATACGCCCTGTCTCTCCCTCAGACCACC 154027  
|||||

QY 939 gagtacctagaatttacctacctctcaaggactgtcccgaggacttccatgaccgacct 998  
|||||  
Db 154028 GAGTGCCTAGAAATTTATGTGCTCTCAAGGACTGTGCCAGGACTTTGTGGCACGGCCTA 154087  
|||||

QY 999 aagacttgcacacagctgttcattcgcgcacatagactgggaaggaggactgtaatt 1058  
|||||  
Db 154088 AAGATTATGCCAACACACTGTTTCATCTGCCGCATTGTGGACTGGAAGGAGGACTGCAATT 154147  
|||||

QY 1059 ttgcctctggggcaactg 1075  
|||||  
Db 154148 TTGCCCTGGGGTAGGTG 154164  
|||||

Search completed: July 8, 2002, 10:07:28  
Job time: 7775 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 08:49:53 ; Search time 329.97 Seconds

(without alignments)  
16130.057 Million cell updates/sec

Title: us-09-647-377-8  
Perfect score: 3100  
Sequence: 1 cggcgccgcgcctcccg.....tgaaaaaaaaaaaaaaaa 3100

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3100	100.0	3100	20 AA223894	Murine LOBO homolo
2	956	30.8	1550	20 AA223889	Murine LOBO cDNA f
3	473	15.3	36901	20 AA223892	Murine LOBO genomi
4	473	15.3	38886	20 AA223897	Murine LOBO homolo
5	236	7.6	49999	20 AA223891	Murine LOBO genomi
6	236	7.6	49999	20 AA223896	Murine LOBO homolo
7	116	3.7	49999	20 AA223895	Murine LOBO homolo
8	49	1.6	925	22 AAK60530	Human Immune/haema
9	49	1.6	1140	20 AA223890	Human LOBO cDNA fr

10	49	1.6	1784	20 AA223898	Human LOBO homolog
11	49	1.6	13330	20 AA223893	Human LOBO genomic
12	49	1.6	49999	20 AA223901	Human LOBO homolog
13	36	1.2	331	21 AAC30796	Human secreted pro
14	32	1.0	428	22 AAS41220	cDNA encoding nove
15	29	0.9	49999	20 AA223903	Human LOBO homolog
16	28	0.9	49999	20 AA223899	Human LOBO homolog
17	26	0.8	1499	22 AAF97934	Human secreted pro
18	25	0.8	513	22 AAF59500	Western flower thr
19	25	0.8	553	22 AAS32943	DNA encoding CARDI
20	25	0.8	2576	22 AAF45119	Human secreted pro
21	25	0.8	2771	21 AAS50743	Rat sensory transd
22	25	0.8	2852	21 AAS58765	cDNA encoding a le
23	25	0.8	2904	19 AAV57904	Bovine butyrophilli
24	24	0.8	223	24 ABK16212	Human lung tumour
25	24	0.8	226	24 ABK16154	Human lung tumour
26	24	0.8	520	22 AAS29115	cDNA encoding for
27	24	0.8	624	21 AAF46469	Partial sequence T
28	24	0.8	793	19 AAV00228	Tick vasoactive am
29	24	0.8	793	20 AAX76965	Histamine binding
30	24	0.8	1422	21 AAC77164	Human ORFX ORF2719
31	24	0.8	1430	23 AAS83088	DNA encoding novel
32	24	0.8	1473	21 AAA39071	Human secreted pro
33	24	0.8	1508	23 AAS83397	DNA encoding novel
34	24	0.8	1723	23 AAS84384	DNA encoding novel
35	24	0.8	2691	21 AAF79685	Pinus radiata cell
36	24	0.8	49999	20 AA223900	Human LOBO homolog
37	23	0.7	123	21 AAA42723	Human secreted exp
38	23	0.7	207	24 ABK15955	Human lung tumour
39	23	0.7	221	24 ABK16135	Human lung tumour
40	23	0.7	221	24 ABK16140	Human lung tumour
41	23	0.7	293	21 AAC98294	Human colon cancer
42	23	0.7	428	19 AAV38430	cDNA encoding the
43	23	0.7	593	22 AAF33107	Human secreted pro
44	23	0.7	712	22 ABA11740	Human nervous syst
45	23	0.7	822	21 AAF21795	Human breast and o

ALIGNMENTS

RESULT 1

AAZ23894

ID AAZ23894 standard; cDNA to mRNA; 3100 BP.

XX AC AAZ23894;

XX DT 25-JAN-2000 (first entry)

XX DE Murine LOBO homologue cDNA to mRNA.

XX KW LOBO; long bones; bone development; bone extension; skull; osteopathic;

XX KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;

XX KW spondyloepiphyseal dysplasia; achondroplasia; murine; ss.

XX OS Mus musculus.

XX Key Location/Qualifiers

FT 125..2737

FT /\*tag= a

FT /product= "LOBO homologue"

XX PN WC9950284-A2.

XX PD 07-OCT-1999.

XX PF 26-MAR-1999; 99WO-EP02055.

XX PR 27-MAR-1998; 98DE-1013799.

XX PA (ROSE/) ROSENTHAL A.

XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;

XX WPI: 1999-601320/51.  
DR P-PSDB; AY50354.  
XX  
XX Nucleic acids encoding proteins which influence bone development,  
PT useful for treating and studying bone disorders -  
XX  
XX Claim 1b; Page 126-130; 391pp; German.  
PS  
XX This invention describes novel nucleic acids (I; designated LOBO (long  
CC bones)) encoding proteins influencing bone development in mammals. The  
CC proteins of the invention reduce and/or inactivate bone extension (i.e.  
CC development), with exception of the skull and have osteopathic activity.  
CC The nucleic acid molecules, proteins and antibodies can be used in  
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
CC and nucleic acid molecules, etc. are useful for production of transgenic  
CC animals, especially a transgenic mouse for the study of diseases  
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and  
CC achondroplasia. This sequence encodes a murine LOBO protein described  
CC in the method of the invention.  
XX  
XX Sequence 3100 BP; 783 A; 795 C; 812 G; 710 T; 0 other;

Query Match 100.0%; Score 3100; DB 20; Length 3100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 Cggcgccgcgcgcctccggagcgagcgtcgtgacaactgagctgtaagcgcgagg 60  
DB 1 Cggcgccgcgcgcctccggagcgagcgtcgtgacaactgagctgtaagcgcgagg 60  
QY 61 aactctgagctgtaagtagtggtccctgaatctggagagagagcacccttggacc 120  
DB 61 aactctgagctgtaagtagtggtccctgaatctggagagagagcacccttggacc 120  
QY 121 agtaataaacatcctgactacaagctaaacttcgcttcctcgaggacccacagagt 180  
DB 121 agtaataaacatcctgactacaagctaaacttcgcttcctcgaggacccacagagt 180  
QY 181 gtctctgtgtgtggccgagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240  
DB 181 gtctctgtgtgtggccgagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240  
QY 241 caagtccatgcgagggagaaagacatatttgaaacctacatgtccaaagagagat 300  
DB 241 caagtccatgcgagggagaaagacatatttgaaacctacatgtccaaagagagat 300  
QY 301 ttcagaaggcttgaaagagaggaacacttatccagggtgtattgagaaatcaacc 360  
DB 301 ttcagaaggcttgaaagagaggaacacttatccagggtgtattgagaaatcaacc 360  
QY 361 gtttcataagccttcattcctctccggatggtgtgtgtgtgtgtgtgtgtgtgtgt 420  
DB 361 gtttcataagccttcattcctctccggatggtgtgtgtgtgtgtgtgtgtgtgtgt 420  
QY 421 tttgtgtcgttaataagagcttaaatggggaccttggtgtgtgtgtgtgtgtgtgtgt 480  
DB 421 tttgtgtcgttaataagagcttaaatggggaccttggtgtgtgtgtgtgtgtgtgtgt 480  
QY 481 tcagtgaagcgaggttaaacaccagagagcaatgacaaagaaatagagctactatga 540  
DB 481 tcagtgaagcgaggttaaacaccagagagcaatgacaaagaaatagagctactatga 540  
QY 541 tgacatccctgaagaggctgtggacatcacccctgcagcagctcccgaaagctggag 600  
DB 541 tgacatccctgaagaggctgtggacatcacccctgcagcagctcccgaaagctggag 600  
QY 601 tggctcgtgattcattatagagctcagtttgatgacagcgactcagaagatagacat 660  
DB 601 tggctcgtgattcattatagagctcagtttgatgacagcgactcagaagatagacat 660  
QY 661 caacaccagtgccgt 720

DB 661 caacaccagtgccgt 720  
QY 721 aaaagaagattctagtactccagttatgaaagatgagaacacccccataccacagac 780  
DB 721 aaaagaagattctagtactccagttatgaaagatgagaacacccccataccacagac 780  
QY 781 aaaggccttatcagagaagtagtcaacttcagaaatcagcaaaagggtgtttacattcttgagaa 840  
DB 781 aaaggccttatcagagaagtagtcaacttcagaaatcagcaaaagggtgtttacattcttgagaa 840  
QY 841 aaagcattctcagagcaaaactggcattcctgaaactcttggtctgataagaacagtagcct 900  
DB 841 aaagcattctcagagcaaaactggcattcctgaaactcttggtctgataagaacagtagcct 900  
QY 901 gtttaagaatacgcctgtttctcttcagaccacagtagtacctagaatttaccatc 960  
DB 901 gtttaagaatacgcctgtttctcttcagaccacagtagtacctagaatttaccatc 960  
QY 961 tctcaaggagctgtcccccaggaacttcagccagcactaaagactttgccaacacgctgt 1020  
DB 961 tctcaaggagctgtcccccaggaacttcagccagcactaaagactttgccaacacgctgt 1020  
QY 1021 catctgcccatcataagtagtgaaggaagtagtgaattttgccttggggcaacttgctcaa 1080  
DB 1021 catctgcccatcataagtagtgaaggaagtagtgaattttgccttggggcaacttgctcaa 1080  
QY 1081 gagtcttgggcagggctgtgaaatcgagcctgaacagaagggatgatactgacagaatatgg 1140  
DB 1081 gagtcttgggcagggctgtgaaatcgagcctgaacagaagggatgatactgacagaatatgg 1140  
QY 1141 tgtggactctctgattctcttcagaaagttctgaatgtctcctcaagcctgcctcg 1200  
DB 1141 tgtggactctctgattctcttcagaaagttctgaatgtctcctcaagcctgcctcg 1200  
QY 1201 gacaaatccacactgatgaggtgggagagaagagacaaactaaaggaaagactgtatctcac 1260  
DB 1201 gacaaatccacactgatgaggtgggagagaagagacaaactaaaggaaagactgtatctcac 1260  
QY 1261 cattgatccatcaactgctcgacacttgaatgatgccctcgctgcaggcggtcactga 1320  
DB 1261 cattgatccatcaactgctcgacacttgaatgatgccctcgctgcaggcggtcactga 1320  
QY 1321 tggcaacttcgaagtggggtccacatcgccagtagtgagttacttttctcctgagggatc 1380  
DB 1321 tggcaacttcgaagtggggtccacatcgccagtagtgagttacttttctcctgagggatc 1380  
QY 1381 ctctttggataaagttagctgtgagagagccacaagtgtctacttggctccagaggtgt 1440  
DB 1381 ctctttggataaagttagctgtgagagagccacaagtgtctacttggctccagaggtgt 1440  
QY 1441 ccccatgtctccaggtctgtgtgaggaactctgagcctcaacccccactgactgacaa 1500  
DB 1441 ccccatgtctccaggtctgtgtgaggaactctgagcctcaacccccactgactgacaa 1500  
QY 1501 gctgacctctctgtgatactggaaagctgaccctgaaaggcaagatccttgaagagtggt 1560  
DB 1501 gctgacctctctgtgatactggaaagctgaccctgaaaggcaagatccttgaagagtggt 1560  
QY 1561 tggcgccactatcactcgttcttgcacaaactgagtaagcactgacacccatgagatgat 1620  
DB 1561 tggcgccactatcactcgttcttgcacaaactgagtaagcactgacacccatgagatgat 1620  
QY 1621 cgaaaatccaaactgagaagatcctgaggaagcttccccaatttctccagagcacag 1680  
DB 1621 cgaaaatccaaactgagaagatcctgaggaagcttccccaatttctccagagcacag 1680  
QY 1681 cgtcagagagtgcaacagcagtagtctgaaactgcaagcattgcaaaagcaactccgcg 1740  
DB 1681 cgtcagagagtgcaacagcagtagtctgaaactgcaagcattgcaaaagcaactccgcg 1740  
QY 1741 ccagcgctttagtagtgcgactcgtttagatcagctgaagcttgcttttactctgga 1800

Db 1741 ccagcgtttgtagatggcgcactccgcttagatcaagctgaagcttgccttttactctgga 1800  
Qy 1801 ccatgagactggactgctcaaggatgttcacatctatgtatccagagacagacaagct 1860  
Db 1801 ccatgagactggactgctcaaggatgttcacatctatgtatccagagacagacaagct 1860  
Qy 1861 tftagagagattcatgctcctggccaaatggcggtggcccaagatcttccgcaactt 1920  
Db 1861 tftagagagattcatgctcctggccaaatggcggtggcccaagatcttccgcaactt 1920  
Qy 1921 cccctgagcggcctgtgctggccggcatccccccacacagacgaagatgctcagtgacct 1980  
Db 1921 cccctgagcggcctgtgctggccggcatccccccacacagacgaagatgctcagtgacct 1980  
Qy 1981 ggttgagattctgtgaccagatggggtggccatggatgctcagctgctgcaggggccctaaa 2040  
Db 1981 ggttgagattctgtgaccagatggggtggccatggatgctcagctgctgcaggggccctaaa 2040  
Qy 2041 taaaacccctgactaagacatttggagatgacagatgctctggccgggaagggtgct 2100  
Db 2041 taaaacccctgactaagacatttggagatgacagatgctctggccgggaagggtgct 2100  
Qy 2101 caccacacatgactcccgcccatgcatgacagatggcactgtacttctgctggtgagtgca 2160  
Db 2101 caccacacatgactcccgcccatgcatgacagatggcactgtacttctgctggtgagtgca 2160  
Qy 2161 ggaccagagcagttccggcattatgtctctcaacgtttccctctcacacacacttcaactc 2220  
Db 2161 ggaccagagcagttccggcattatgtctctcaacgtttccctctcacacacacttcaactc 2220  
Qy 2221 tcccatccgcgtttgtgacagtcatagtgacccgctcctggtgctgctgctgggcta 2280  
Db 2221 tcccatccgcgtttgtgacagtcatagtgacccgctcctggtgctgctgctgggcta 2280  
Qy 2281 cagtgaacacccagatgtggagcctgtatccctacagaagcaagctgacacactgcaatga 2340  
Db 2281 cagtgaacacccagatgtggagcctgtatccctacagaagcaagctgacacactgcaatga 2340  
Qy 2341 ccgtcgcatggtctccaaacgtgtgcagagagctcagcatcgccctctcttcgcagttct 2400  
Db 2341 ccgtcgcatggtctccaaacgtgtgcagagagctcagcatcgccctctcttcgcagttct 2400  
Qy 2401 agtaaaagagatggccctctggagtcgaagccatgtgtatgggtgctcctgaaaccaagc 2460  
Db 2401 agtaaaagagatggccctctggagtcgaagccatgtgtatgggtgctcctgaaaccaagc 2460  
Qy 2461 ttctgacgtgctgtgctgctgttgggtgacagaagcgcactactgcaatgcaactggc 2520  
Db 2461 ttctgacgtgctgtgctgctgttgggtgacagaagcgcactactgcaatgcaactggc 2520  
Qy 2521 cctgcgcatctacagcttccagaagtgagggaagcaagcagctcaactcttctgttgga 2580  
Db 2521 cctgcgcatctacagcttccagaagtgagggaagcaagcagctcaactcttctgttgga 2580  
Qy 2581 gcctgagtaccttgaagagagcccaacacagcagctcatcacctctcagcctgtgga 2640  
Db 2581 gcctgagtaccttgaagagagcccaacacagcagctcatcacctctcagcctgtgga 2640  
Qy 2641 tctggtcctgcaggcagagccacagccctcaagtacagtgctatctctgaaagcagaag 2700  
Db 2641 tctggtcctgcaggcagagccacagccctcaagtacagtgctatctctgaaagcagaag 2700  
Qy 2701 ccttggaagagcgtctgagtggagcctgaggaactgaatgctagcccaagcaggcctgt 2760  
Db 2701 ccttggaagagcgtctgagtggagcctgaggaactgaatgctagcccaagcaggcctgt 2760  
Qy 2761 gcctgcctacccctgctgcttttaggaataaggaccttttgacacaaagggatttta 2820  
Db 2761 gcctgcctacccctgctgcttttaggaataaggaccttttgacacaaagggatttta 2820  
Qy 2821 atttggttttaaacactcagggttctgtttttatttttttttttttttttttttttactt 2880  
Db 2821 atttggttttaaacactcagggttctgtttttatttttttttttttttttttttttactt 2880

Qy 2881 ttgcagctcaagcttttaataactgaactggaaggttaggggtcagggcagggatgctgaggc 2940  
Db 2881 ttgcagctcaagcttttaataactgaactggaaggttaggggtcagggcagggatgctgaggc 2940  
Qy 2941 ctggcctgtgcttccctgagcagagagatccagctcctcctgggcagcagcccgctt 3000  
Db 2941 ctggcctgtgcttccctgagcagagagatccagctcctcctgggcagcagcccgctt 3000  
Qy 3001 ctaccagcagccacactgcccctccctgccagaaaaatgggggttttcagcaaatcagtg 3060  
Db 3001 ctaccagcagccacactgcccctccctgccagaaaaatgggggttttcagcaaatcagtg 3060  
Qy 3061 tcatggaataaatcaagtgtgaaaaaataaaaaa 3100  
Db 3061 tcatggaataaatcaagtgtgaaaaaataaaaaa 3100

## RESULT 2

AAZ23889  
ID AAZ23889 standard; cDNA; 1550 BP.

XX AC AAZ23889;

XX XX 25-JAN-2000 (first entry)

XX XX Murine LOBO cDNA fragment.

XX XX LOBO; long bones; bone development; bone extension; skull; osteopathic;

XX KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;

XX KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.

XX OS Mus musculus.

XX PH Key Location/Qualifiers

XX FT CDS I..1183

XX FT /\*tag= a

XX FT /product= "LOBO"

XX XX WO950284-A2.

XX XX 07-OCT-1999.

XX XX 26-MAR-1999; 99WO-EP02055.

XX XX 27-MAR-1998; 98DE-1013799.

XX XX (ROSE/) ROSENTHAL A.

XX XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;

XX XX WPI; 1999-601320/51.

XX XX P-PSDB; AAY50352.

XX XX Nucleic acids encoding proteins which influence bone development,

XX XX useful for treating and studying bone disorders -

XX XX Example 3; Page 62-64; 391pp; German.

XX XX This invention describes novel nucleic acids (I; designated LOBO (long

XX XX bones)) encoding proteins influencing bone development in mammals. The

XX XX proteins of the invention reduce and/or inactivate bone extension (i.e.

XX XX development), with exception of the skull and have osteopathic activity.

XX XX The nucleic acid molecules, proteins and antibodies can be used in

XX XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods

XX XX and nucleic acid molecules, etc. are useful for production of transgenic

XX XX animals, especially a transgenic mouse for the study of diseases

XX XX associated with bone development, e.g. spondyloepiphyseal dysplasia and

XX XX achondroplasia. This sequence encodes the murine LOBO protein described

XX XX in the method of the invention.

XX XX Sequence 1550 BP; 352 A; 433 C; 412 G; 353 T; 0 other;

```
Query Match      30.8%; Score 956; DB 20; Length 1550;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2129 atggcaactgactctgtctggtggtgctgagcagaccaggagcaggttccgcgattatgct 2188
      |||||||
Db 575 atggcaactgactctgtctggtggtgctgagcagaccaggagcaggttccgcgattatgct 634
      |||||||
QY 2189 ctcaagcttccctctacacacacttcaaccttcccatccgccttctgtcagcgtcaata 2248
      |||||||
Db 635 ctcaagcttccctctacacacacttcaaccttcccatccgccttctgtcagcgtcaata 694
      |||||||
QY 2249 gtgcacccgctcctggtgtgctgtctggtggtggtggtggtggtggtggtggtggtggt 2308
      |||||||
Db 695 gtgcacccgctcctggtgtgctgtggtggtggtggtggtggtggtggtggtggtggtggt 754
      |||||||
QY 2309 accctacagaagaagcgtgacacactgcaatgacccgtgcgactggttcccaaacgtgtgcaq 2368
      |||||||
Db 755 accctacagaagaagcgtgacacactgcaatgacccgtgcgactggttcccaaacgtgtgcaq 814
      |||||||
QY 2369 gagctacagcagcctctcttctgagcttctgagcttctgagcttctgagcttctgagctt 2428
      |||||||
Db 815 gagctacagcagcctctcttctgagcttctgagcttctgagcttctgagcttctgagcttct 874
      |||||||
QY 2429 gaagccatggtgagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 2488
      |||||||
Db 875 gaagccatggtgagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 934
      |||||||
QY 2489 gtgcagaagcgtcactactgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 2548
      |||||||
Db 935 gtgcagaagcgtcactactgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 994
      |||||||
QY 2549 ggggaagaagcagagctcactctgttgggagcctgagcctgagcctgagcctgagcctgagc 2608
      |||||||
Db 995 ggggaagaagcagagctcactctgttgggagcctgagcctgagcctgagcctgagcctgagc 1054
      |||||||
QY 2609 cagcaggtcactcaccatctcagcctggtggtggtggtggtggtggtggtggtggtggtggt 2668
      |||||||
Db 1055 cagcaggtcactcaccatctcagcctggtggtggtggtggtggtggtggtggtggtggtggt 1114
      |||||||
QY 2669 ctcaagtaagtgctatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2728
      |||||||
Db 1115 ctcaagtaagtgctatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1174
      |||||||
QY 2729 gaggactgaatgctagcccaagccagcctgctgctgctgctgctgctgctgctgctgctgctg 2788
      |||||||
Db 1175 gaggactgaatgctagcccaagccagcctgctgctgctgctgctgctgctgctgctgctgctg 1234
      |||||||
QY 2789 ataggaccttttgacacaaaggggatttttaattggttttaacaaactcaggggtttg 2848
      |||||||
Db 1235 ataggaccttttgacacaaaggggatttttaattggttttaacaaactcaggggtttg 1294
      |||||||
QY 2849 tttttattttatttttcccttttatttttacttttgcagctcagtttttaaatgaaactgga 2908
      |||||||
Db 1295 tttttattttatttttcccttttatttttacttttgcagctcagtttttaaatgaaactgga 1354
      |||||||
QY 2909 aggttaggggtcagggcaggggagctgctgagggcctggtgctgctgctgctgctgctgctg 2968
      |||||||
Db 1355 aggttaggggtcagggcaggggagctgctgagggcctggtgctgctgctgctgctgctgctg 1414
      |||||||
QY 2969 atccacagtcctcctggtgagcagcagcccgcttctaccagggaccacactgcccctccctg 3028
      |||||||
Db 1415 atccacagtcctcctggtgagcagcagcccgcttctaccagggaccacactgcccctccctg 1474
      |||||||
QY 3029 ccacaggaatgggggttttcagcaaatcagtgctcagtgctcagtgctcagtgctcagtgct 3084
      |||||||
Db 1475 ccacaggaatgggggttttcagcaaatcagtgctcagtgctcagtgctcagtgctcagtgct 1530
      |||||||

RESULT 3
AAZ23892
ID AAZ23892 standard; DNA; 36901 BP.
```

```
XX AAZ23892;
AC
XX
XX 25-JAN-2000 (first entry)
DT
XX
XX Murine LOBO genomic DNA fragment 2.
DE
XX
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmacological; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
OS Mus musculus.
XX
XX WO9950284-A2.
PN
XX
XX 07-OCT-1999.
PD
XX
XX 26-MAR-1999; 99WO-EP02055.
PF
XX
XX 27-MAR-1998; 98DE-1013799.
PR
XX
XX (ROSE/) ROSENTHAL A.
PA
XX
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
PI
XX
XX WPI; 1999-601320/51.
DR
XX
XX Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
XX Example 3; Page 97-117; 391pp; German.
XX
XX This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
XX Sequence 36901 BP; 9597 A; 9409 C; 8973 G; 8922 T; 0 other;
SQ

Query Match      15.3%; Score 473; DB 20; Length 36901;
Best Local Similarity 100.0%; Pred. No. 4.5e-169;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2612 caggtcatcaccatctcagcctgtgtggtggtggtggtggtggtggtggtggtggtggtggt 2671
      |||||||
Db 18080 caggtcatcaccatctcagcctgtgtggtggtggtggtggtggtggtggtggtggtggtggt 18139
      |||||||
QY 2672 aagtacagtgtctatcctgaagcaccagcctggtggtggtggtggtggtggtggtggtggtggt 2731
      |||||||
Db 18140 aagtacagtgtctatcctgaagcaccagcctggtggtggtggtggtggtggtggtggtggtggt 18199
      |||||||
QY 2732 gactgaatctgagcccaagcagcctgctgctgctgctgctgctgctgctgctgctgctgctg 2791
      |||||||
Db 18200 gactgaatctgagcccaagcagcctgctgctgctgctgctgctgctgctgctgctgctgctg 18259
      |||||||
QY 2792 ggacctttgacacaaagggttttaattggttttaacaaactcaggggtttgttt 2851
      |||||||
Db 18260 ggacctttgacacaaagggttttaattggttttaacaaactcaggggtttgttt 18319
      |||||||
QY 2852 ttattttatttttcttttatttttacttttgcagctcagtttttaaatgaaactggaagg 2911
      |||||||
Db 18320 ttattttatttttcttttatttttacttttgcagctcagtttttaaatgaaactggaagg 18379
      |||||||
QY 2912 ttagggttcagggcaggggagctgagggcctgctgctgctgctgctgctgctgctgctgctgctg 2971
      |||||||
```

```
Db 18380 ttagggggtcaggccaggggatgctgtaggccttgcttgcttccctcctgagcagagagatc 18439
Qy 2972 ccagtcctctcggcagcagcccgcttctaccagggaaccactgcccttccctgccc 3031
Db 18440 ccagtcctctcggcagcagcccgcttctaccagggaaccactgcccttccctgccc 18499
Qy 3032 aggaatgggggttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 3084
Db 18500 aggaatgggggttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 18552

RESULT 4
ID AAZ23897
XX AAZ23897 standard; DNA; 38886 BP.
AC AAZ23897;
XX 25-JAN-2000 (first entry)
XX Murine LOBO homologue genomic DNA fragment 3.
DE LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX Mus musculus.
XX WO9950284-A2.
XX 07-OCT-1999.
XX 26-MAR-1999; 99WO-EP02055.
XX 27-MAR-1998; 98DE-1013799.
XX (ROSE/) ROSENTHAL A.
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI; 1999-601320/51.
XX Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX Example 3; Page 189-211; 391pp; German.
XX This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX Sequence 38886 BP; 10087 A; 9920 C; 9356 G; 9523 T; 0 other;

Query Match 15.3%; Score 473; DB 20; Length 38886;
Best Local Similarity 100.0%; Pred. No. 4.5e-169;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2612 caggtcatccatctcagccttggtgatgtgtcctcagcagagccacagccctc 2671
Db 20064 caggtcatccatctcagccttggtgatgtgtcctcagcagagccacagccctc 20123
Qy 2672 aagtcagtgtctatcctgagcagccagcctgaggaagcgtctgatgagagcctgag 2731
Db 20124 aagtcagtgtctatcctgagcagccagcctgaggaagcgtctgatgagagcctgag 20183
```

```
Qy 2732 gactgaatgtagcccaagccagcagcctgtgctcctaccctgctggcttttaggaata 2791
Db 20184 gactgaatgtagcccaagccagcagcctgtgctcctaccctgctggcttttaggaata 20243
Qy 2792 ggaccttttagacaccaaaagggtattttaatttggttttaacaactcagggggtttgtt 2851
Db 20244 ggaccttttagacaccaaaagggtattttaatttggttttaacaactcagggggtttgtt 20303
Qy 2852 ttattttatttttcttttattttacttttgcagtcagtcagtttttaaatgaactggaagg 2911
Db 20304 ttattttatttttcttttattttacttttgcagtcagtcagtttttaaatgaactggaagg 20363
Qy 2912 ttagggggtcagggcagggaatgctgagccttgccctgtgcttccctgagcagagagatc 2971
Db 20364 ttagggggtcagggcagggaatgctgagccttgccctgtgcttccctgagcagagagatc 20423
Qy 2972 ccagtcctcctgagcagcagcccgcttctaccaggcgaccactgcccttccctgccc 3031
Db 20424 ccagtcctcctgagcagcagcccgcttctaccaggcgaccactgcccttccctgccc 20483
Qy 3032 aggaatgggggtttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 3084
Db 20484 aggaatgggggtttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 20536

RESULT 5
AAZ23891
ID AAZ23891 standard; DNA; 49999 BP.
XX AAZ23891;
XX 25-JAN-2000 (first entry)
XX Murine LOBO genomic DNA fragment 1.
DE LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX Mus musculus.
XX WO9950284-A2.
XX 07-OCT-1999.
XX 26-MAR-1999; 99WO-EP02055.
XX 27-MAR-1998; 98DE-1013799.
XX (ROSE/) ROSENTHAL A.
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI; 1999-601320/51.
XX Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX Example 3; Page 69-97; 391pp; German.
XX This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;
```

```
Query Match      7.6%; Score 236; DB 20; Length 49999;
Best Local Similarity 100.0%; Pred. No. 9.5e-80;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1542 agatcctgaagagtggtttggcgcactatcatcgttcttgaccacaaactgagctacg 1601
      |||
Db 8518 agatcctgaagagtggtttggcgcactatcatcgttcttgaccacaaactgagctacg 8577

QY 1602 accatgccagacatgatcgaataatccaaactgagaagatccctgaggaagagcttcccc 1661
      |||
Db 8578 accatgccagacatgatcgaataatccaaactgagaagatccctgaggaagagcttcccc 8637

QY 1662 caatttctccagagcagcgtcgagaggtgcaccaggcagtcctgaaacctgcacagca 1721
      |||
Db 8638 caatttctccagagcagcgtcgagaggtgcaccaggcagtcctgaaacctgcacagca 8697

QY 1722 ttgcaaaagaactccgcgcagcgtttgttagatggcgcaactccgttttagatcag 1777
      |||
Db 8698 ttgcaaaagaactccgcgcagcgtttgttagatggcgcaactccgttttagatcag 8753

RESULT 6
AAZ23896
ID AAZ23896 standard; DNA; 49999 BP.
AC AAZ23896;
XX
XX
DT 25-JAN-2000 (first entry)
DE Murine LOBO homologue genomic DNA fragment 2.
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
XX
OS Mus musculus.
XX
XX PN WO9950284-A2.
XX PD 07-OCT-1999.
XX PF 26-MAR-1999; 99WO-EP02055.
XX PR 27-MAR-1998; 98DE-1013799.
XX PA (ROSE/) ROSENTHAL A.
XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX DR WPI; 1999-601320/51.
XX PT Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX PS Example 3; Page 161-189; 391pp; German.
XX CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
XX SQ Sequence 49999 BP; 13135 A; 11787 C; 10868 G; 14209 T; 0 other;

Query Match      3.7%; Score 116; DB 20; Length 49999;
Best Local Similarity 100.0%; Pred. No. 1.6e-34;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1542 agatcctgaagagtggtttggcgcactatcatcgttcttgaccacaaactgagctacg 1601
      |||
Db 10503 agatcctgaagagtggtttggcgcactatcatcgttcttgaccacaaactgagctacg 10562

QY 1602 accatgccagacatgatcgaataatccaaactgagaagatccctgaggaagagcttcccc 1661
      |||
Db 10563 accatgccagacatgatcgaataatccaaactgagaagatccctgaggaagagcttcccc 10622

QY 1662 caatttctccagagcagcgtcgagaggtgcaccaggcagtcctgaaacctgcacagca 1721
      |||
Db 10623 caatttctccagagcagcgtcgagaggtgcaccaggcagtcctgaaacctgcacagca 10682

QY 1722 ttgcaaaagaactccgcgcagcgtttgttagatggcgcaactccgttttagatcag 1777
      |||
Db 10683 ttgcaaaagaactccgcgcagcgtttgttagatggcgcaactccgttttagatcag 10738

RESULT 7
AAZ23895
ID AAZ23895 standard; DNA; 49999 BP.
AC AAZ23895;
XX
XX
DT 25-JAN-2000 (first entry)
DE Murine LOBO homologue genomic DNA fragment 1.
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
XX OS Mus musculus.
XX
XX PN WO9950284-A2.
XX PD 07-OCT-1999.
XX PF 26-MAR-1999; 99WO-EP02055.
XX PR 27-MAR-1998; 98DE-1013799.
XX PA (ROSE/) ROSENTHAL A.
XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX DR WPI; 1999-601320/51.
XX PT Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX PS Example 3; Page 133-161; 391pp; German.
XX CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
XX SQ Sequence 49999 BP; 14550 A; 10214 C; 10719 G; 14516 T; 0 other;
```











—  
—  
—  
—

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 8, 2002, 07:57:53 ; Search time 329.03 Seconds  
(without alignments)  
16176.138 Million cell updates/sec

Title: US-09-647-377-8  
Perfect score: 3100  
Sequence: 1 cggcgccgcgcctcccg.....tgaaaaaaaaaaaaaaaaaaaaa 3100

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues 3472872

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3100	100.0	3100	20 AAZ23894	Murine LOBO homolo
2	1330.6	42.9	1550	20 AAZ23889	Murine LOBO CDNA f
3	1181.8	38.1	1784	20 AAZ23898	Human LOBO homolo
4	565.6	18.2	1140	20 AAZ23890	Human LOBO CDNA fr
5	473	15.3	36901	20 AAZ23892	Murine LOBO genomi
6	473	15.3	38886	20 AAZ23897	Murine LOBO homolo
7	277.8	9.0	925	22 AAK60530	Human immune/haema
8	236.8	7.6	49599	20 AAZ23891	Murine LOBO genomi
9	236.8	7.6	49599	20 AAZ23896	Murine LOBO homolo

10	235.4	7.6	331	21 AAC30796	Human secreted pro
11	203	6.5	428	22 AAC31220	cDNA encoding nove
12	187.6	6.1	5347	23 ABL11058	Drosophila melanog
13	183.4	5.9	2955	23 ABL02003	Drosophila melanog
14	183.4	5.9	5373	23 ABL02002	Drosophila melanog
15	182.2	5.9	13330	20 AAZ23893	Human LOBO homolo
16	182.2	5.9	49999	20 AAZ23901	Human LOBO homolo
17	180.8	5.8	3119	23 ABL11059	Drosophila melanog
18	176	5.7	49999	20 AAZ23899	Human LOBO homolo
19	150.4	4.9	49999	20 AAZ23900	Human LOBO homolo
20	116	3.7	49999	20 AAZ23895	Murine LOBO homolo
21	114.2	3.7	49999	20 AAZ23903	Human LOBO homolo
22	96	3.1	11309	19 AAV52241	Streptococcus pneu
23	95.6	3.1	2379	22 AAH53049	S. epidermidis ope
24	95.6	3.1	3111	22 AAH54223	S. epidermidis gen
25	94.2	3.0	702	21 AAF14474	Aspergillus oryzae
26	91.4	2.9	2354	21 AAA05471	Streptococcus pneu
27	85	2.7	273254	21 AAC81914	Chlamydia pneumoni
28	76.4	2.5	6005	20 AAX13264	Enterococcus faeca
29	75.4	2.4	2944528	24 ABA03041	Listeria monocytog
30	69.2	2.2	11185	22 AAS28662	Genomic sequence #
31	68.2	2.2	1744	22 AAS41202	cDNA encoding nove
32	68	2.2	468	22 AAL36030	Human musculoskele
33	68	2.2	468	22 AAL36031	Human musculoskele
34	68	2.2	96109	22 AAF28548	Genomic fragment #
35	64.6	2.1	123	21 AAA42723	Human secreted exp
36	64.6	2.0	6157	18 AAF74495	Staphylococcus aur
37	62.4	2.0	5191	21 AAF79601	Virulence gene #21
38	60.2	1.9	1007	21 AAF07819	Fusarium venenatum
39	60	1.9	5146	23 ABL11080	Drosophila melanog
40	59.8	1.9	3584	22 AAS28661	Drosophila melanog
41	59.8	1.9	4309	23 AAS79116	Genomic sequence #
42	55.2	1.8	2376	21 AAZ54518	DNA encoding novel
43	55.2	1.8	2391	24 AAI67439	Neisseria meningit
44	54	1.7	373	22 AAS41212	N. meningitidis va
45	50	1.6	1926	21 AAZ54517	cDNA encoding nove
					Neisseria meningit

ALIGNMENTS

RESULT 1  
NAZ23894  
ID AAZ23894 standard; cDNA to mRNA; 3100 BP.

AC AAZ23894;  
XX  
XX 25-JAN-2000 (first entry)  
DT  
DE Murine LOBO homologue cDNA to mRNA.

XX LOBO; long bones; bone development; bone extension; skull; osteopathic;  
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ss.

OS Mus musculus.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 125..2737  
FT /\*tag= a  
FT /product= "LOBO homologue"

XX WO9950284-A2.  
XX  
XX 07-OCT-1999.  
XX  
XX 26-MAR-1999; 99WO-EP02055.  
XX  
XX 27-MAR-1998; 98DE-1013799.  
XX (ROSE/) ROSENTHAL A.  
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;  
PI



QY	2881	ttgcagctcagtttttaaatgaactggaagttaagggtcagggcagggcaggaatgctgaggc	2940
Db	2881	ttgcagctcagtttttaaatgaactggaagttaagggtcagggcagggcaggaatgctgaggc	2940
QY	2941	ctggcctgtgttcctcagcagagagatcccagtcctctgagcaggcagcccgctt	3000
Db	2941	ctggcctgtgttcctcagcagagagatcccagtcctctgagcaggcagcccgctt	3000
QY	3001	ctaccagcgaccactgcccctccctcccgccagagaatgggggtttcagcaaatcagtg	3060
Db	3001	ctaccagcgaccactgcccctccctcccgccagagaatgggggtttcagcaaatcagtg	3060
QY	3061	tcatggataaaaatcaagctgtgaaaaaataaaaaaaaaa 3100	
Db	3061	tcatggataaaaatcaagctgtgaaaaaataaaaaaaaaa 3100	
RESULT 2			
ID	AAZ23889 standard; cDNA; 1550 BP.		
XX	AAZ23889;		
AC	AAZ23889;		
DT	25-JAN-2000 (first entry)		
XX	Murine LOBO cDNA fragment.		
DE			
XX			
KW	LOBO; long bones; bone development; bone extension; skull; osteopathi		
KW	diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;		
KW	spondyloepiphyseal dysplasia; achondroplasia; murine; ds.		
OS	Mus musculus.		
XX	Key	Location/Qualifiers	
FH	1..1183		
FT	CDS	/*tag= a	
FT		/product= "LOBO"	
XX	W0950284-A2.		
PN			
PD	07-OCT-1999.		
XX			
XX	26-MAR-1999;	99WO-EP02055.	
XX			
XX	27-MAR-1998;	98DE-1013799.	
XX			
PA	(ROSE/) ROSENTHAL A.		
PI	Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;		
XX			
XX	WPI: 1999-601320/51.		
DR	P-PSDB; AAY50352.		
XX			
XX	Nucleic acids encoding proteins which influence bone development,		
FT	useful for treating and studying bone disorders -		
XX			
PS	Example 3; Page 62-64; 391pp; German.		
XX			
CC	This invention describes novel nucleic acids (I; designated LOBO (long		
CC	bones)) encoding proteins influencing bone development in mammals. The		
CC	proteins of the invention reduce and/or inactivate bone extension (i.e.		
CC	development), with exception of the skull and have osteopathic activi		
CC	The nucleic acid molecules, proteins and antibodies can be used in		
CC	diagnostic or pharmaceutical compounds e.g. for gene therapy. The met		
CC	and nucleic acid molecules, etc. are useful for production of transgen		
CC	animals, especially a transgenic mouse for the study of diseases		
CC	associated with bone development, e.g. spondyloepiphyseal dysplasia an		
CC	achondroplasia. This sequence encodes the murine LOBO protein describ		
CC	in the method of the invention.		

Query Match 42.9%; Score 1330.6; DB 20; Length 1550;  
Best Local Similarity 89.4%; Pred. No. 0;  
Matches 1530; Conservative 0; Mismatches 4; Indels 177; Gaps 2;

QY 1390 taaagttagctgtgagagagccacaagtgtgtctacttgggtccagaaagtgtgtccccatgct 1449  
Db 13 taaagttagctgtgagagagccacaagtgtgtctacttgggtccagaaagtgtgtcccatgct 72

QY 1450 tccagagcttctgtgtgagaaactctgcagctcaaccccatgaactgacagagctgacctt 1509  
Db 73 tccagagcttctgtgtgagaaactctgcagctcaaccccatgaactgacagagctgacctt 132

QY 1510 ctctgtgatctggaagctgaacctgaagcgaagatccttgaagagtggttggccgcac 1569  
Db 133 ctctgtgatctggaagctgacctgaagcgaagatccttgaagagtggttggccgcac 192

QY 1570 tatcatcgttctgcacaaactgagctacgacctgcccagagcagcatgacgaataac 1629  
Db 193 tatcatcgttctgcacaaactgagctacgacctgcccagagcagcatgacgaataac 252

QY 1630 aactgagaagatccctggaagagcttcccccatttctccagagcagcgtcgagga 1689  
Db 253 aactgagaagatccctggaagagcttcccccatttctccagagcagcgtcgagga 312

QY 1690 ggtgcaccaggagctgctgaacctgcacagcattgcaaaagcaactccgcgcagcgctt 1749  
Db 313 ggtgcaccaggagctgctgaacctgcacagcattgcaaaagcaactccgcgcagcgctt 372

QY 1750 tgttagatggcgcaactccgttttagatcagctgaagctgtgttactctggaccatgagac 1809  
Db 373 tgttagatggcgcaactccgttttagatc----- 398

QY 1810 tggactgctcaagatgtcacatctatgagtaccgagacagcaacaagctgttagagga 1869  
Db 399 -----agga 402

QY 1870 gtctactgctctggccaactggcggtggccacaagatcttccgcacacttccctgagca 1929  
Db 403 gtctactgctctggccaactggcggtggccacaagatcttccgcacacttccctgagca 462

QY 1930 ggcctctgctgcgcggcgtcccccacacagacgaagatgctcagtgacctgggtgagtt 1989  
Db 463 ggcctctgctgcgcggcgtcccccacacagacgaagatgctcagtgacctgggtgagtt 522

QY 1990 ctgtgaccagatggggctgcccagtgatgctcagctctgcaggggccctaaataaagcct 2049  
Db 523 ctgtgaccagatggggctgcccagtgatgctcagctctgcaggggccctaaataa----- 574

QY 2050 gactaagacatttgagatgacaagtactctgtggccgggaaggaggtgctcaccaacat 2109  
Db 575 ----- 574

QY 2110 gtactccggcccatgcagatggcactgtactctgtctgagtgctgcagagaccagga 2169  
Db 575 -----atggcactgtactctgtctctggtggaatgctgcagagaccagga 615

QY 2170 geagttccggcaattatgctctcaactgtccctctacacacacttcaactctcccatccg 2229  
Db 616 geagttccggcaattatgctctcaactgtccctctacacacacttcaactctcccatccg 675

QY 2230 ccgcttctgctgacgtcatagtgaaccgctcctggctgctgctgctgaggtacagtaagca 2289  
Db 676 ccgcttctgctgacgtcatagtgaaccgctcctggctgctgctgctgaggtacagtaagca 735

QY 2290 gccagatgtggagctgataccctacagaagcaagctgaccactgcaatgacctgcgat 2349  
Db 736 gccagatgtggagctgataccctacagaagcaagctgaccactgcaatgacctgcgat 795

QY 2350 ggcctccaaagctgtgcagagctcagcatcgacctcttcttcagcagttctagtaagga 2409  
Db 796 ggcctccaaagctgtgcagagctcagcatcgacctcttcttcagcagttctagtaagga 855

QY 2410 gagtggcccttgagtcgagtcgagccatggtgatggtgtgctgacccaagctttcgagct 2469  
Db 856 gagtggcccttgagtcgagtcgagccatggtgatggtgtgctgacccaagctttcgagct 915

QY 2470 gctggtgctgcgctttgggtgcagaagcgcatctactgcaatgcactgacagcctgcgac 2529  
Db 916 gctggtgctgcgctttgggtgcagaagcgcatctactgcaatgcactgacagcctgcgac 975

QY 2530 ctacagcttccagaaggtggggaagccagagctcactcttgtttggagcctgcagta 2589  
Db 976 ctacagcttccagaaggtggggaagccagagctcactcttgtttggagcctgcagta 1035

QY 2590 ccttgaagagagcgaacacacagcagctcatcacatcttcagcctgctgagatgagctt 2649  
Db 1036 ccttgaagagagcgaacacacagcagctcatcacatcttcagcctgctgagatgagctt 1095

QY 2650 ggagcagagggccacagccctcaagtacagtactctctctgaagcagcagcctggagaa 2709  
Db 1096 ggagcagagggccacagccctcaagtacagtactctctgaagcagcagcctggagaa 1155

QY 2710 ggcgtctgatgagagcctgagagctgaatgctagcccaagccagcctgtgctgcacct 2769  
Db 1156 ggcgtctgatgagagcctgagagctgaatgctagcccaagccagcctgtgctgcacct 1215

QY 2770 acctgctgctttagaagtaggaccttttgacacaaaggaggatttttaatttggttt 2829  
Db 1216 acctgctgctttagaagtaggaccttttgacacaaaggaggatttttaatttggttt 1275

QY 2830 ttacaactcagggttggtt 2889  
Db 1276 ttacaactcagggttggtt 1335

QY 2890 agtttttaaaatgaactggaaggttaggggtcagggcagggagatgcagagcctgacct 2949  
Db 1336 agtttttaaaatgaactggaaggttaggggtcagggcagggagatgcagagcctgacct 1395

QY 2950 gcttcctgagcagagagatccagctcctcctggcagcagcccgcttctaccagcc 3009  
Db 1396 gcttcctgagcagagagatccagctcctcctggcagcagcccgcttctaccagcc 1455

QY 3010 gacccactgccttccctccaggaagtgggggttttcagcaaatcagtgatcaggaat 3069  
Db 1456 gacccactgccttccctccaggaagtgggggttttcagcaaatcagtgatcaggaat 1515

QY 3070 aaatcaagtgtgaaaaaataaaaaa 3100  
Db 1516 aaatcaagtgtgaaaaaataaaaaa 1546

RESULT 3  
AAZ23898  
ID AAZ23898 standard; CDNA to mRNA; 1784 BP.  
XX  
AC AAZ23898;  
XX  
DT 25-JAN-2000 (first entry)  
XX  
DE Human LOBO homologue cDNA to mRNA.  
XX  
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;  
diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
spondyloepiphyseal dysplasia; achondroplasia; human; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 1-1454  
FT /\*tag= a  
FT /product= "LOBO homologue"  
FT /note= "Partial sequence, no start codon given"  
XX  
PN W09950284-A2.  
XX

PD 07-OCT-1999.  
XX  
PF 26-MAR-1999; 99WO-EP02055.  
XX  
PR 27-MAR-1998; 98DE-1013799.  
XX  
PA (ROSE/) ROSENTHAL A.  
XX  
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;  
XX WPI; 1999-601320/51.  
DR P-PSDB; AAY50355.  
XX  
PT Nucleic acids encoding proteins which influence bone development,  
PT useful for treating and studying bone disorders -  
XX  
PS Claim 1b; Page 211-213; 391pp; German.  
XX  
CC This invention describes novel nucleic acids (I; designated LOBO (long  
CC bones)) encoding proteins influencing bone development in mammals. The  
CC proteins of the invention reduce and/or inactivate bone extension (i.e.  
CC development), with exception of the skull and have osteopathic activity.  
CC The nucleic acid molecules, proteins and antibodies can be used in  
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
CC and nucleic acid molecules, etc. are useful for production of transgenic  
CC animals, especially a transgenic mouse for the study of diseases  
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and  
CC achondroplasia. This sequence encodes a human LOBO protein described  
CC in the method of the invention.  
XX  
SQ Sequence 1784 BP; 385 A; 537 C; 504 G; 358 T; 0 other;

Query Match 38.1%; Score 1181.8; DB 20; Length 1784;  
Best Local Similarity 81.9%; Pred. No. 2.4e-301;  
Matches 1471; Conservative 0; Mismatches 267; Indels 59; Gaps 7;

QY 1323 gcaccttcgaagtggcgccacatcgccgagtgagttactttgttcctgaggatcct 1382  
DB 1 gcaacttcaaaagtgagggttcacattgctgagtgagttactttgttcctgaggatcgt 60  
QY 1383 ctttgataaagttagctgtgagagccagacaaagtctacttgggtccagaagtggtccc 1442  
DB 61 atccggataaagtggcgccgagagggctacaaagcgtctacttgggtccaaaggggtgccc 120  
QY 1443 ccattgctccaggcttctgtgaggaaactctgcagcctcaaccccatgactgacaagc 1502  
DB 121 ccattgctccaggcttctgtgaggagctgtgcagcctcaaccccatgctccgacaagc 180  
QY 1503 tgacctctctgtgactggaagctgacccctgaaggcaagatcccttgagagtggtttg 1562  
DB 181 tgacctctctgtgactggaagctgacccctgaaggagggcgaagatcccttgatgaatggtttg 240  
QY 1563 gccgcactatcactcgtttctgcaccaactgagctacgacctgcccagagcatgatgacg 1622  
DB 241 gccgcacatcactcgtctctgcaccaacttagctacgagcatgcaagagagatgattg 300  
QY 1623 aaaaatccaactgaaagatccctgaggaagagcttcccctaattctccagagcacagcg 1682  
DB 301 aaagcccaactgaaataacccctgcgaagagagctgccccatttcccagagacatagca 360  
QY 1683 tcgaggaggtgacacaggcagtcctgaacctgcacagcatgcaaaagcaactccgcgcgc 1742  
DB 361 gcgaggaggtacacaggccgcttgaatctccacggattgccaagcagttacgcccagc 420  
QY 1743 agcgcttgtagtgcgcactcgttttagatcacgtgaagcttgctttactctggacc 1802  
DB 421 agcgcttgtagtgcgcactcgttttagatcacgtgaagcttgctttactctggacc 480  
QY 1803 atgagactggagctcctcaaggatgtcaatcatatgataccgagacacaaagcttg 1862  
DB 481 acgagaccggattgctcgaaggatgcatatcatatgataccgagagacaagaactcg 540

QY 1863 tagagagattcatctcctgcccacacatggcggtggcccccacagatcttccgcacttcc 1922  
DB 541 tggaggagttcatctctctgtggccaaatggcgagtggtggcccaagatccacgcgccttcc 600  
QY 1923 ctgagcagggcctgctgcgcggcatccccccacacagacagaaagtctcagtgacctgg 1982  
DB 601 ccgagcagggcctgctgcgcggcaccccccgcccccaacaaggatgctcagtgacctgg 660  
QY 1983 tggagttctgacacagatgggactcccatgagtgctcagctgaggggcccctaaata 2042  
DB 661 tggaaattctgaccagatgggctggcctggacttccagctccgcagggagccctcaata 720  
QY 2043 aaagcctgactaaagacatttggagatgacaagtactctctggcccggaaaggagtgctca 2102  
DB 721 aaagcctgaccacaaacatttggagatgacaagtactctgctggcggcgaaggagtgctca 780  
QY 2103 ccaaatgactacccggcccatgacagatgacatgacttctctctgggagtgctgcagg 2162  
DB 781 ccaaatgctgctcccgcccatgacagatgacatgacttctctctgggagtgctgcagg 840  
QY 2163 accagagcagttccggcattatgctctcaacggttccctctacacacattcacctctc 2222  
DB 841 accagcgcagttccggcactacgcgtcaatgtgccctgtacacacattcacctctc 900  
QY 2223 caatccgcgcttctgagctgacgtacatgtgacgcgctcctgctgctgctgggtctaca 2282  
DB 901 caatccgcgcttctgagcgtcctgggtgacgcgctcctgctgctgctgggtctata 960  
QY 2283 gtgaacagccagatgtggagcctgataccctacagaagcgaagcgtgacacatgcaatgacc 2342  
DB 961 gggagcagtagacatggcgccgataccctgcagaacagcggaacctgtaacgacc 1020  
QY 2343 gtgcgactgtccaaacgctgacagagctcagcatcgccctcttcttcgacgttctag 2402  
DB 1021 gccgatggctccaaagcgctgacagagctcagtcacagctctcttctgtgcttgg 1080  
QY 2403 taaagagagtggtggccctgagtgccgaagcagatggtgagtggttctcctgaacacagctt 2462  
DB 1081 tcaagagagtggtggccctgagtgcaagacccatggtgagtggtggtccatccctgaagcagcct 1140  
QY 2463 tcgacgtgctggtgctgcgcttgggtgcagaagcgcgcatctactgcaatgcaactggccc 2522  
DB 1141 tcgacgtgctggtgctgcgctacggtgcgagcgtgcagaagcgcgcatctactgcaacgcaactggccc 1200  
QY 2523 tgcgactctacagcttcagaagtggtgggaagcagacagctcactctgtttgtggagc 2582  
DB 1201 tgcggtcccaactccagaaagtggggaagcagcagcagcagcagcagcagcagcagc 1260  
QY 2583 ctgatgaccttgaagagggccacacagcagcagcagcagcagcagcagcagcagcagcagc 2642  
DB 1261 ctgagagcatggagcagggcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320  
QY 2643 tggctcctgagcagagggccacagcagcagcagcagcagcagcagcagcagcagcagcagc 2702  
DB 1321 tggctcctgagcagagggccacagcagcagcagcagcagcagcagcagcagcagcagcagc 1380  
QY 2703 t-----ggagaagcgtctgtaggagcgtgagagcgtgagagcgtgagagcgtgagagcgt 2735  
DB 1381 cccaggggcaacctgggcccctgagaagggaggaggagcgtgacggtgagcggcagagcgt 1440  
QY 2736 gaa-----tgctagcccacagcagcgtgctgcccctgacgtgagcggcagagcgt 2782  
DB 1441 caagcagcagcgtgagctccacagcagcagcagcagcagcagcagcagcagcagcagcagc 1500  
QY 2783 ---ttagaataaggacctttgacacaaaggagatttttaatttggtttttaaacactc 2839  
DB 1501 aactggtctttagacctgtgacacgaggggggttttaatttggtttttaaacactc 1560  
QY 2840 aggggttggtt 2899  
DB 1561 aggggttggtt 1608  
QY 2900 tgaactggaaggttaggggtcagg--gcaggggagtgctgagggcctggcctgtgcttccct 2957





```
SQ Sequence 38886 BP; 10087 A; 9920 C; 9356 G; 9523 T; 0 other;

Query Match          15.3%; Score 473; DB 20; Length 38886;
Best Local Similarity 100.0%; Pred. No. 1.7e-113;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2612 caggtcatcaccatttcagcctggtggtgctgctcagggcagagccacagccctc 2671
      |||
Db 20064 caggtcatcaccatttcagcctggtggtgctgctcagggcagagccacagccctc 20123

QY 2672 aagtacagtgctatcctgaagcaccagcctgagaagcgtctgatgagagcctgag 2731
      |||
Db 20124 aagtacagtgctatcctgaagcaccagcctgagaagcgtctgatgagagcctgag 20183

QY 2732 gactgaatgtagcccaagccagcctgtgctgcctaccctgctgcttttaggaata 2791
      |||
Db 20184 gactgaatgtagcccaagccagcctgtgctgcctaccctgctgcttttaggaata 20243

QY 2792 ggacctttgacacaaagggttttaatttggttttaacaaactcaggggtttgtt 2851
      |||
Db 20244 ggacctttgacacaaagggttttaatttggttttaacaaactcaggggtttgtt 20303

QY 2852 ttatttttttctctttattttacttttgcagctcagtttttaaatgaactggaag 2911
      |||
Db 20304 ttatttttttctctttattttacttttgcagctcagtttttaaatgaactggaag 20363

QY 2912 ttagggttcagggcagggatgctgagcctgcctgtgcttcctcagcagagagatc 2971
      |||
Db 20364 ttagggttcagggcagggatgctgagcctgcctgtgcttcctcagcagagagatc 20423

QY 2972 ccagtcctcctgggcagcagcccgcttaccaggcgaccactgcccttcctgcgcc 3031
      |||
Db 20424 ccagtcctcctgggcagcagcccgcttaccaggcgaccactgcccttcctgcgcc 20483

QY 3032 aggaatgggggttttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 3084
      |||
Db 20484 aggaatgggggttttcagcaaatcagtgctcatggaataaaatcaagtgtgaa 20536

RESULT 7
AAK60530
ID AAK60530 standard; cDNA; 925 BP.
XX
AC AAK60530;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:5590.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 07-JUN-2000; 2000US-0205515.
PR 28-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0214886.
PR 07-JUL-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
```



```

XX Murine LOBO genomic DNA fragment 1.
DE
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
XX Mus musculus.
OS
XX WO9950284-A2.
PN
XX 07-OCT-1999.
PD
XX 26-MAR-1999; 99WO-EF02055.
PF
XX 27-MAR-1998; 98DE-1013799.
PR
XX (ROSE/) ROSENTHAL A.
PA
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
PI WPI; 1999-601320/51.
DR
XX Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
XX Example 3; Page 69-97; 391pp; German.
PS
XX This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;

Query Match 7.6%; Score 236.8; DB 20; Length 49999;
Best Local Similarity 99.2%; Pred. No. 4.4e-51;
Matches 238; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1542 agatcctgaagagtggttggccgcaactatcctcgttcttgaccacaaactgagctacg 1601
DB 8518 agatcctgaagagtggttggccgcaactatcctcgttcttgaccacaaactgagctacg 1601
QY 1602 accatgccagagcatgatcgaatacctgagaagatccctgagaagagcttcccc 1661
DB 8578 accatgccagagcatgatcgaatacctgagaagatccctgagaagagcttcccc 1661
QY 1662 caattctccagagcaagctcgaggaggtgcaccaggagctcctgaacctgcacagca 1721
DB 8638 caattctccagagcaagctcgaggaggtgcaccaggagctcctgaacctgcacagca 1697
QY 1722 ttgcaaaagcaactccgcgcagcgttttagatggcgacactccgttttagatcagctga 1781
DB 8698 ttgcaaaagcaactccgcgcagcgttttagatggcgacactccgttttagatcagctga 8757

RESULT 9
AAZ23896
ID AAZ23896 standard; DNA; 49999 BP.
XX
AC AAZ23896;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO homologue genomic DNA fragment 2.

XX Murine LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
XX Mus musculus.
OS
XX WO9950284-A2.
PN
XX 07-OCT-1999.
PD
XX 26-MAR-1999; 99WO-EF02055.
PF
XX 27-MAR-1998; 98DE-1013799.
PR
XX (ROSE/) ROSENTHAL A.
PA
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
PI WPI; 1999-601320/51.
DR
XX Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
XX Example 3; Page 69-97; 391pp; German.
PS
XX This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13135 A; 11787 C; 10868 G; 14209 T; 0 other;

Query Match 7.6%; Score 236.8; DB 20; Length 49999;
Best Local Similarity 99.2%; Pred. No. 4.4e-51;
Matches 238; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1542 agatcctgaagagtggttggccgcaactatcctcgttcttgaccacaaactgagctacg 1601
DB 10503 agatcctgaagagtggttggccgcaactatcctcgttcttgaccacaaactgagctacg 10562
QY 1602 accatgccagagcatgatcgaatacctgagaagatccctgagaagagcttcccc 1661
DB 10563 accatgccagagcatgatcgaatacctgagaagatccctgagaagagcttcccc 10622
QY 1662 caattctccagagcaagctcgaggaggtgcaccaggagctcctgaacctgcacagca 1721
DB 10623 caattctccagagcaagctcgaggaggtgcaccaggagctcctgaacctgcacagca 10682
QY 1722 ttgcaaaagcaactccgcgcagcgttttagatggcgacactccgttttagatcagctga 1781
DB 10683 ttgcaaaagcaactccgcgcagcgttttagatggcgacactccgttttagatcagctga 10742

RESULT 10
AAC30796
ID AAC30796 standard; cDNA; 331 BP.
XX
AC AAC30796;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 34871.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

```



PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI P-PSDB; AAU23350.  
XX  
XX Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT Pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases -  
XX  
XX Claim 4; SEQ ID No 446; 1180pp; English.  
XX  
XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAS40785-AAS41684 represent cDNA sequences encoding for the novel human  
CC enzyme polypeptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 428 BP; 101 A; 94 C; 122 G; 106 T; 5 other;

Query Match 6.5%; Score 203; DB 22; Length 428;  
Best Local Similarity 87.4%; Pred. No. 2.7e-43;  
Matches 221; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

QY 1067 gggcaactggcctaagagctgtggcaggctggtgaaatcagcctgaaacacagaggata 1126  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||  
Db 1 gggcagctggcctaagagctgtggcaggctggtgaaatcagcctgaaacacagaggata 60  
QY 1127 ctgcacagaatagtgtggactctctgtattctcttcacagaagttctgaatgtctcct 1186  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||  
Db 61 ctaacagagatagcgtggtattctctgtattctcttcacagaagttctgaatgtctcct 120  
QY 1187 caagcctgcccctggacaatccacctatgatggcggcgaagagacacgaagaaa 1246  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||  
Db 121 caagcctgcccctggacaatccacctatgatggcggcgaagagacacgaagaaa 180  
QY 1247 gactgtatctccacttgcattcattcattcattcattcattcattcattcattcattc 1306  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||  
Db 181 gactgtatctccacttgcattcattcattcattcattcattcattcattcattcattc 240  
QY 1307 aggcgctcactg 1319  
||| ||||| |||  
Db 241 aagcactcactg 253

RESULT 12











Align seg 1/1 to: AA223894 from: 1 to: 3100

```
1 MetAsnHisProAspTyrLysLeuAsnLeuArgSerProGlyThrProAr 17
|||||
125 ATGAACCATCTGACTACAAAGCTGAACCTCGGTCTCCGGGACCCCACT 174
|||||
17 gGlyValSerSerValValGlyProSerAlaValGlyAlaSerProGlyA 34
|||||
175 AGGTGTGTCCTCTGTGTTGGCCGAGTGTCTGTTGGTCTTCGCGCAGGTG 224
|||||
34 sPLysLysSerLysAsnLysSerMetArgGlyLysLysLysSerIlePhe 50
|||||
225 ACAAAAAGTCAAGAACAAAGTCCATCGGAGGGAAGAAAAGAGCATATT 274
|||||
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67
|||||
275 GAAACCTACATGTCTCAAGGAGGATGTTTCAGAAAGCTTGAAGAGAGGAAC 324
|||||
67 rLeuIleGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84
|||||
325 ACTTATCCAGGGTGTATTGAGATCAACCCAAAGAGTTTCATGAGCCCT 374
|||||
84 heIleProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100
|||||
375 TCATTCTCTCTCCGGATGGTGTATCGGGACATTTTATTGATGGAGTTGTT 424
|||||
101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLysLeuLeuPr 117
|||||
425 GCTCGTANTAGAGCCTTAATGGGACCTTGTGGTTGTAAACCTTGCCTCC 474
|||||
117 oGluAspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIleG 134
|||||
475 TGAGGATCAGTGGGAAGCGAGTTAAACACAGAGCAATGACAAAAGAAATAG 524
|||||
134 luAlaThrTyrGluAlaAspIleProGluGluGlyCysGlyHisHisPro 150
|||||
525 AAGCTACTTATGAGCTGACATCCCTGAAGAGGCTGTGGACATCACCCC 574
|||||
151 LeuGlnGlnSerArgLysGlyTrpSerGlyProAspValIleIleGluAl 167
|||||
575 CTGCAGAGAGTCCCGAAAGGCTGGAGTGGTCTGATGCTATTATAGAGC 624
|||||
167 acIlnPheAspAspSerAspSerGluAspArgHisGlyAsnThrSerGlyL 184
|||||
625 TCAGTTTGATGACAGCGACTCAGAAGATAGACATGGCAACACCAAGTGCC 674
|||||
184 euValAspGlyValLysLysLeuSerIleSerThrProAspArgGlyLys 200
|||||
675 TGGTTGATGGTGTAAAGAAATGTCAATCTCTACTCTGACAGAGGAAA 724
|||||
201 GluAspSerSerThrProValMetLysAspGluAsnThrProIleProgl 217
|||||
725 GAAGATTCTAGTACTCCAGTTATGAAGATGAGAAACACCCCATCCACA 774
|||||
217 nAspThrArgGlyLeuSerGluLysSerLeuGlnLysSerAlaLysValV 234
|||||
775 GGACACAAGAGGCTTATCAGAGAGTCACITTCAGAAATCAGCAAGAGTGG 824
|||||
234 alTyrIleLeuGluLysLysHisSerArgAlaIleThrGlyIleLeuLys 250
|||||
825 TTTACATCTTGGGAAAAAGCAATTCGAGCAGCAACTGGCACTCCTGAAA 874
|||||
251 LeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaLeuPheSe 267
|||||
875 CTCTTGGCTGATPANGAACAGTACCCTGTTTAAAGAAATACGCCCTGTTTC 924
|||||
267 rProSerAspHisArgValProArgIleTyrValProLeuLysAspCysP 284
|||||
925 TCCTTCAGACCCAGTAGTACCTAGAAATTTAGCTACCTCAAGGACTGTC 974
|||||
284 roGlnAspPheMetThrArgProLysAspPheAlaAsnThrLeuPheIle 300
|||||
975 CCCAGGACTTCATGACCCGACCTTAAGACTTTGCCAACACAGCGTGTTCATC 1024
|||||
```

```
301 CysArgIleIleAspTrpLysGluAspCysAsnPheAlaLeuGlyGlnLe 317
|||||
1025 TGGCGCATCATAGACTGGGAAGGAGACTGTAATTTTGCCTCGGGCAACT 1074
|||||
317 uAlaLysSerLeuGlyGlnAlaGlyGluIleGluProGluThrGluGlyI 334
|||||
1075 GGCTAAGAGTCTTGGCGAGGCTGGTGAATTCGAGCCTGAACAGAGGGA 1124
|||||
334 leLeuThrGluTyrGlyValAspPheSerAspPheSerSerGluValLeu 350
|||||
1125 TACTGACAGAAATATGTTGGACTTCTCTGATTTCTTTCAGAAAGTTCTT 1174
|||||
351 GluCysLeuProGlnSerLeuProTrpThrIleProProAspGluValG 367
|||||
1175 GAATGTCTCCCTCAAAAGCCTGCCCTGGACAAATCCACCTGATGAGGTGG 1224
|||||
367 yLysArgArgAspLeuArgLysAspCysIlePheThrIleAspProSerT 384
|||||
1225 CAAGAGAAAGAGACCTAAGGAAAGACTGTATCTTCACCATTTGATCCATCA 1274
|||||
384 hrAlaArgAspLeuAspAspAlaLeuAlaCysArgArgLeuThrAspGly 400
|||||
1275 CTGCTCGCACCTTGATGATGCCCTCGCCTGCAGCGGCTCACTGATGGC 1324
|||||
401 ThrPheGluValGlyValHisIleAlaAspValSerTyrPheValProG 417
|||||
1325 ACCTTCGAAAGTGGGCTCCACATCGCGATGTGAGTTACTTTGTTCCTGA 1374
|||||
417 uGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrL 434
|||||
1375 GGGATCCTCTTTGGATAAAGTAGTGTGTGAGAGAGCCACAAGTGTCTACT 1424
|||||
434 euValGlnLysValValProMetLeuProArgLeuLeuCysGluGluLeu 450
|||||
1425 TGTCTCAGAGAGGTGTCTCCCATGCTTCCCAGGCTTCTGTGTGAGGAAGCTC 1474
|||||
451 CysSerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLy 467
|||||
1475 TGCAGCCTCAACCCCATGACTGACAAAGCTGACCTTCTGTGTGATCTGGAA 1524
|||||
467 sLeuThrProGluGlyLysIleLeuGluGluTrpPheGlyArgThrIleI 484
|||||
1525 GCTGACCCCTGAGCGCAGATCCTTGAAGAGTGGTTTGGCCGCACTATCA 1574
|||||
484 leArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGlu 500
|||||
1575 TCGGTTCTTGACCAAACTGAGCTACGACCATGCCAGCATGATCGAA 1624
|||||
501 AsnProThrGluLysIleProGluGluGluLeuProProIleSerProGl 517
|||||
1625 AATCCAACCTGAGAGATCCCTGAGGAGAGCTTCCCCCAATTTCTCCAGA 1674
|||||
517 uHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleA 534
|||||
1675 GCACAGCGCTCGAGGAGGTGCACAGGACAGTCTCTGAACCTGCACAGCATTG 1724
|||||
534 laLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAasp 550
|||||
1725 CAAAGCAACTCCGCGCCAGCGCTTTGTAGATGGCGCACTCCGTTTAGAT 1774
|||||
551 GlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnGl 567
|||||
1775 CAGCTGAAAGCTTGTCTTTTACTCTGGACCATGAGACTGGACTGCCCTCAAGG 1824
|||||
567 yCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPheM 584
|||||
1825 ATGTCACATCTATGAGTACCGAGACAGCAACAAGCTTGTAGAGAGTTCA 1874
|||||
584 etLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhePro 600
|||||
1875 TGCTCTCTGGCCAAACATGGCGGTGGCCCAACAAATCTTCGCACACCTCCCT 1924
|||||
```

601 GluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSe 617  
1925 GAGCAGGCCCTGCTGGCGGGCATCCGCCACAGAGATGCTCAG 1974  
617 rAspLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSerS 634  
1975 TGACCTGGTGGAGTTCGTGACAGATGGGGCTGCCATGGATGTCAGCT 2024  
634 eAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspLys 650  
2025 CTCGAGGGGCCCTAAATAAAGCCTGACTAAGACATTTGGAGATGACAAG 2074  
651 TyrSerLeuAlaArgLysGlnValLeuThrAsnMetTyrSerArgProMe 667  
2075 TACTCTGGCCCGAAGAGGTGCTACCCACATGTACTCCGGCCCAT 2124  
667 tGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGlnP 684  
2125 GCAGATGGCATGCTACTTCTGCTCTGGGATGCTGCAGGACGAGGACGT 2174  
684 heArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerPro 700  
2175 TCCGGCATTATGCTCTCAAGGTTCCCTCTACACACACTTCACCTCTCCC 2224  
701 IleArgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaIle 717  
2225 ATCCGCCGCTTTGGTGCAGTCAATAGTGCACCGCTCCCTGGCTGCTCT 2274  
717 uGlyTyrSerGluGlnProAspValGluProAspThrLeuGlnLysGlnA 734  
2275 GGGCTACAGTGAACAGCCATGCTGGAGCCTGATACCTACAGAGCAAG 2324  
734 laAspHisCysAsnAspArgArgMetAlaSerLysArgValGlnGlnLeu 750  
2325 CTGACACCTCAATGACCGTCGATGGCTTCCAAACGTGTGCAGAGGCTC 2374  
751 SerIleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuG1 767  
2375 AGCATCGCCCTCTTCTTCGAGTTCCTAGTAAAGGAGAGTGCCCCCTGGA 2424  
767 uSerGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuV 784  
2425 GTCCGAAGCAATGGTATGGGTGCTCTGAACCAAGCTTTCACCTGTGG 2474  
784 alLeuArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeu 800  
2475 TGTCTGGCTTTGGGGTGCAGAAGCGCATCTACTGCAATGCACCTGGCCCTG 2524  
801 ArgSerTyrSerPheGlnLysValGlyLysProGluLeuThrLeuVa 817  
2525 CGATCTACAGCTTCCAGAGGTGGGGAAGAGCCAGAGCTCCTCTTGT 2574  
817 lTrpGluProAspAspLeuGluGluProThrGlnGlnValIleThrI 834  
2575 TTGGGAGCTGATGACCTTGAAGAGGAGCAACACAGCAGGTTCATACCA 2624  
834 lPheSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLys 850  
2625 TCTTCAGCTGGTGGATGGTCTCTGCAGGCGGAGGCCACAGCCCTCAAG 2674  
851 TyrSerAlaIleLeuLysArgProGlyLeuGluLysAlaSerAspGluG1 867  
2675 TACAGTGTCTATCCTGAAGCGACCGGCTGGAGAAAGGGCTCTGTATGAGGA 2724  
867 uProGluAsp 870  
2725 GCCTGAGGAC 2734

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.AAZ23889

seq\_documentation\_block:  
ID\_AAZ23889 standard; cDNA; 1550 BP.  
XX

AAZ23889;  
25-JAN-2000 (first entry)  
Murine LOBO cDNA fragment.  
LOBO; long bones; bone development; bone extension; skull; osteopathic;  
diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
spondyloepiphyseal dysplasia; achondroplasia; murine; ds.  
Mus musculus.  
Key Location/Qualifiers  
CDS 1..1183  
/\*tag= a  
/product= "LOBO"  
WO950284-A2.  
07-OCT-1999.  
26-MAR-1999; 99WO-EP02055.  
27-MAR-1998; 98DE-1013799.  
(ROSE/) ROSENTHAL A.  
Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;  
WPI; 1999-601320/51.  
P-PSDB; AAY50352.  
Nucleic acids encoding proteins which influence bone development,  
useful for treating and studying bone disorders -  
Example 3; Page 62-64; 391pp; German.  
This invention describes novel nucleic acids (I; designated LOBO (long  
bones)) encoding proteins influencing bone development in mammals. The  
proteins of the invention reduce and/or inactivate bone extension (i.e.  
development), with exception of the skull and have osteopathic activity.  
The nucleic acid molecules, proteins and antibodies can be used in  
diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
and nucleic acid molecules, etc. are useful for production of transgenic  
animals, especially a transgenic mouse for the study of diseases  
associated with bone development, e.g. spondyloepiphyseal dysplasia and  
achondroplasia. This sequence encodes the murine LOBO protein described  
in the method of the invention.  
Sequence 1550 BP; 352 A; 433 C; 412 G; 353 T; 0 other;  
alignment\_scores:  
Quality: 202.00 Length: 202  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-647-377-9 x AAZ23889 ..  
Align seg 1/1 to: AAZ23889 from: 1 to: 1550  
669 MetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluGlnPheAr 685  
575 ATGGCACTGACTTCTGCTGGATGCTGCAGGACCGAGGAGCAGTTCGG 624  
685 gHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIleA 702  
625 GCATTATGCTCTCAAGTTCCTCTACACACTTCACCTCTCCCATCC 674  
702 rgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGly 718  
675 GCCGCTTTGCTGACGTACATAGTCACCGCTCTCTGGCTGCTGCTGGGC 724

719 TyrSerGluGlnProAspValGluProAspThrLeuGlnLysGlnAlaAs 735  
|||||  
725 TACAGTGAACAGCAGATGTGGAGCTGATACCTACAGAAGCAAGCTGA 774  
735 pHisCysAsnAspArgArgMetAlaSerLysArgValGlnGluSerI 752  
|||||  
775 CCACCTGCAATGACCGTGGATGGCTTCCAAACGTGTGCAGAGCTCAGCA 824  
752 LeGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuGluSer 768  
|||||  
825 TCGGCTCTCTTCGAGTCTTACTAAGAGAGAGTGGCCCCCTGGAGTCC 874  
769 GluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuValLe 785  
|||||  
875 GAAGCATTGGTGTGTGGTGTCTTGAACCAAGCTTTTCGACGTGCTGTGCT 924  
785 uArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArgS 802  
|||||  
925 GCGCTTTGGGTGCAGAGCGCATCTACTGCAATGCATGGCCCTCGCAT 974  
802 erTyrSerPheGlnLysValGlyLysLysProGluLeuThrLeuValTrp 818  
|||||  
975 CCTACAGCTTCCAAGAGGTGGGGAAGAGCCAGAGCTCCTCTTGTGG 1024  
819 GluProAspAspLeuGluGluProThrGlnGlnValIleThrIlePhe 835  
|||||  
1025 GAGCCTGATGACCTTGAAGAGGAGCCAAACAGCAGGTCATCACCATT 1074  
835 eSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLysTyrS 852  
|||||  
1075 CAGCCTGGTGGATGTGGTCTGCAGGCAGAGGCCACAGCCCTCAAGTACA 1124  
852 erAlaIleLeuLysArgProGlyLeuGluLysAlaSerAspGluGluPro 868  
|||||  
1125 GTGCTATCTTGAAGCAGCAGCCCTGGAGAGGCGTCTGATGAGGAGCT 1174  
869 GluAsp 870  
|||||  
1175 GAGGAC 1180

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23891  
seq\_documentation\_block:  
ID AAZ23891 standard; DNA; 49999 BP.  
XX AC AAZ23891;  
XX DT 25-JAN-2000 (first entry)  
XX DE Murine LOBO genomic DNA fragment 1.  
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;  
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.  
XX OS Mus musculus.  
XX PN W09950284-A2.  
XX PD 07-OCT-1999.  
XX PF 26-MAR-1999; 99WO-EP02055.  
XX PR 27-MAR-1998; 98DE-1013799.  
XX PA (ROSE/) ROSENTHAL A.  
XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;  
XX WPI; 1999-601320/51.  
XX DR  
XX PT Nucleic acids encoding proteins which influence bone development,

PT useful for treating and studying bone disorders -  
XX Example 3; Page 69-97; 391pp; German.  
XX This invention describes novel nucleic acids (I; designated LOBO (long bones)) encoding proteins influencing bone development in mammals. The proteins of the invention reduce and/or inactivate bone extension (i.e. development), with exception of the skull and have osteopathic activity. The nucleic acid molecules, proteins and antibodies can be used in diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods and nucleic acid molecules, etc. are useful for production of transgenic animals, especially a transgenic mouse for the study of diseases associated with bone development, e.g. spondyloepiphyseal dysplasia and achondroplasia. This sequence encodes the murine LOBO protein described in the method of the invention.  
XX Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;  
SQ  
alignment\_scores:  
Quality: 78.00 Length: 78  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-647-377-9 x AAZ23891  
Align seg 1/1 to: AAZ23891 from: 1 to: 49999  
474 IleLeuGluGluTrpPheGlyArgThrIleIleArgSerCysThrLysLe 490  
8520 ATCTTTGAAGAGTGGTTTGGCGCAGTATATCATCTCTTGCACCAAACT 8569  
490 uSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysIleP 507  
8570 GAGCTACGACCATGCCAGAGCATGATCGAAATCCAACTGAGAAGATCC 8619  
507 roGluGluGluLeuProProIleSerProGluHisSerValGluGluVal 523  
8620 CTGAGGAAGAGCTTCCCCCAATTTCTCCAGAGCAGCAGCGTCGAGGAGGTG 8669  
524 HisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArgG1 540  
8670 CACGAGGAGTCCCGAACCTGCACAGCATTCGAAAGCAACTCCGCCGCCA 8719  
540 nArgPheValAspGlyAlaLeuArgLeuAspGln 551  
8720 GCGCTTTGTAGATGGCGCACTCGGTTTAGATCAG 8753  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23896  
seq\_documentation\_block:  
ID AAZ23896 standard; DNA; 49999 BP.  
XX AC AAZ23896;  
XX DT 25-JAN-2000 (first entry)  
XX DE Murine LOBO homologue genomic DNA fragment 2.  
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;  
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.  
XX OS Mus musculus.  
XX PN W09950284-A2.  
XX PD 07-OCT-1999.  
XX PF 26-MAR-1999; 99WO-EP02055.  
XX PR 27-MAR-1998; 98DE-1013799.

```
XX (ROSE/) ROSENTHAL A.
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI; 1999-601320/51.
XX Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX Example 3; Page 161-189; 391pp; German.
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes the murine LOBO protein described
XX in the method of the invention.
XX Sequence 49999 BP; 13135 A; 11787 C; 10868 G; 14209 T; 0 other;

alignment_scores:
  Quality: 78.00      Length: 78
  Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23896 ..

Align seg 1/1 to: AAZ23896 from: 1 to: 49999
474 IleLeuGluGluTrpPheGlyArgThrIleIleArgSerCysThrLysLe 490
|||||
10505 ATCTTGAAGAGTGTTGGCGCACATCATCGCTTTCACCAAACT 10554
|||||
490 uSerTyAspHisAlaGlnSerMetIleGluAsnProThrGluLysIleP 507
|||||
10555 GAGTACGACATGCCAGAGCATGATCGAAATCCAACTGAGAAGATCC 10604
|||||
507 roGluGluGluLeuProProIleSerProGluHisSerValGluGluVal 523
|||||
10605 CTGAGGAGAGCTTCCCCAATTTCTCCAGAGCACACGCTCGAGAGGTG 10654
|||||
524 HisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArgI 540
|||||
10655 CACGAGCAGTCTCTGAACCTGCACAGCATTCGAAAGCAACTCGCGCGCA 10704
|||||
540 nArgPheValAspGlyAlaLeuArgLeuAspGln 551
|||||
10705 GCGCTTTGTAGATGGCGCACTCCGTTTAGATCAG_10738

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ23892
seq_documentation_block:
ID AAZ23892 standard; DNA; 36901 BP.
XX AAZ23892;
XX
XX 25-JAN-2000 (first entry)
XX Murine LOBO genomic DNA fragment 2.
XX
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;
XX diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
XX spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
XX Mus musculus.
```

```
XX WO9950284-A2.
XX
XX 07-OCT-1999.
XX
XX 26-MAR-1999; 99WO-BF02055.
XX
XX 27-MAR-1998; 98DE-1013799.
XX
XX (ROSE/) ROSENTHAL A.
XX
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI; 1999-601320/51.
XX Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX Example 3; Page 97-117; 391pp; German.
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes the murine LOBO protein described
XX in the method of the invention.
XX Sequence 36901 BP; 9597 A; 9409 C; 8973 G; 8922 T; 0 other;

alignment_scores:
  Quality: 61.00      Length: 61
  Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23892 ..

Align seg 1/1 to: AAZ23892 from: 1 to: 36901
579 LeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysI 595
|||||
12478 CTTGTAGAGAGGTTCATGCTCTGCGCCACATGGCGTGGCCACAGAT 12527
|||||
595 ePheArgThrPheProGluGlnAlaLeuLeuArgArgHisProProG 612
|||||
12528 CTTCCGCACCTTCCCTGAGCAGCGCTGCTGCGCGGCATCCGCCACC 12577
|||||
612 InThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeu 628
|||||
12578 AGACGAGAGATGCTCAGTCCCTGGTGGAGTTCGTGACCATGATGGG 12627
|||||
629 ProMetAspValSerSerAlaGlyAlaLeuAsn 639
|||||
12628 CCCATGGATGTCAGCTCTGCAGGGGCCCTAAAT 12660
|||||

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ23897
seq_documentation_block:
ID AAZ23897 standard; DNA; 38886 BP.
XX
XX AAZ23897;
XX
XX 25-JAN-2000 (first entry)
XX Murine LOBO homologue genomic DNA fragment 3.
XX
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;
```

```
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
OS Mus musculus.
XX WO9950284-A2.
XX 07-OCT-1999.
XX 26-MAR-1999; 99WO-EP02055.
XX 27-MAR-1998; 98DE-1013799.
XX (ROSE/) ROSENTHAL A.
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI; 1999-601320/51.
XX Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX Example 3; Page 189-211; 391pp; German.
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes the murine LOBO protein described
XX in the method of the invention.
XX Sequence 38886 BP; 10087 A; 9920 C; 9356 G; 9523 T; 0 other;

alignment_scores:
  Quality: 61.00 Length: 61
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23897 ..
Align seg 1/1 to: AAZ23897 from: 1 to: 38886

579 LeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysI 595
|||||
14462 CTTGTAGAGGAGTTTCATGCTCTCGGCAACATGGCGTGGCCCAAGAT 14511

595 ePheArgThrPheProGluGlnAlaLeuLeuArgHisProProG 612
|||||
14512 CTTCCGACACCTCCCTCGAGAGGCCCTGCTGCGCGGCGCATCCGCCACCAC 14561

612 InrLysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeu 628
|||||
14562 AGACGAAGATGCTCAGTACCTGGTGGAGTCTGTGCCAGATGGGGCTG 14611

629 ProMetAspValSerSerAlaGlyAlaLeuAsn 639
|||||
14612 CCCATGATGTCAGTCTGTCAGGGGCCCTAAAT 14644

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ23895
seq_documentation_block:
ID AAZ23895 standard; DNA; 49999 BP.
XX
AC AAZ23895;
XX
DT 25-JAN-2000 (first entry)
```

```
XX Murine LOBO homologue genomic DNA fragment 1.
XX
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
XX Mus musculus.
XX WO9950284-A2.
XX 07-OCT-1999.
XX 26-MAR-1999; 99WO-EP02055.
XX 27-MAR-1998; 98DE-1013799.
XX (ROSE/) ROSENTHAL A.
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI; 1999-601320/51.
XX Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX Example 3; Page 133-161; 391pp; German.
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes the murine LOBO protein described
XX in the method of the invention.
XX Sequence 49999 BP; 14550 A; 10214 C; 10719 G; 14516 T; 0 other;

alignment_scores:
  Quality: 39.00 Length: 39
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23895 ..
Align seg 1/1 to: AAZ23895 from: 1 to: 49999

400 GlyThrPheGluValGlyValHisIleAlaAspValSerTyrPheValPr 416
|||||
30110 GGCACCTTCGAAGTGGCGGTCCACATCGCGCATGTGAGTTACTTTGTTCC 30159

416 oGluGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValIT 433
|||||
30160 TCAGGGATCTCTTTGGATTAAGTAGTCTGCTGAGAGAGCCACAGGTGCT 30209

433 YrLeuValGlnLysVal 438
|||||
30210 ACTTGTCCAGAGGTA 30226

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ23898
seq_documentation_block:
ID AAZ23898 standard; cDNA to mRNA; 1784 BP.
XX
AC AAZ23898;
XX
DT 25-JAN-2000 (first entry)
```

```
XX Human LOBO homologue cDNA to mRNA.
DE
XX LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1454
FT /tag= a
FT /product= "LOBO homologue"
FT /note= "Partial sequence, no start codon given"
XX
PN W09950284-A2.
XX
XX 07-OCT-1999.
XX
XX 26-MAR-1999; 99WO-EP02055.
XX
XX 27-MAR-1998; 98DE-1013799.
XX
XX (ROSE/) ROSENTHAL A.
XX
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
XX WPI; 1999-601320/51.
XX
XX P-PSDB; AAY50355.
XX
XX Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX
XX Claim 1b; Page 211-213; 391pp; German.
XX
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes a human LOBO protein described
XX in the method of the invention.
XX
XX Sequence 1784 BP; 385 A; 537 C; 504 G; 358 T; 0 other;
SQ

alignment_scores:
    Quality: 36.00      Length: 36
    Ratio: 1.000      Gaps: 0
    Percent similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23898 ..
Align seg 1/1 to: AAZ23898 from: 1 to: 1784

421 LeuAspLysValAlaAlaGluArgAlaThrSerValTyrLeuValGlnIly 437
|||||
63 CTGGATAAAGTGGCTGCCGAGAGGGCTACAAAGCGTCTACTTGGTTCAAAA 112
|||||
437 sValValProMetLeuProArgLeuLeuCysGluGluLeuCysSerLeuA 454
|||||
113 GTGTGTCCTCCCATGCTCTCCAGCGTCTGTGTGAGGAGCTGTGCACGCTCA 162
|||||
454 snProMet 456
|||||
163 ACCCCCATG 170

seq_name: /SID51/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAS41220
```

```
seq_documentation_block:
ID_ AAS41220 standard; cDNA; 428 BP.
XX
XX AAS41220;
AC
XX
XX 17-DEC-2001 (first entry)
DT
XX
XX cDNA encoding novel human enzyme polypeptide #436.
DE
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179055.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 11-JUL-2000; 2000US-0217496.
XX
XX 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.
XX
XX 14-AUG-2000; 2000US-0224519.
XX
XX 14-AUG-2000; 2000US-0225213.
XX
XX 14-AUG-2000; 2000US-0225214.
XX
XX 14-AUG-2000; 2000US-0225266.
XX
XX 14-AUG-2000; 2000US-0225267.
XX
XX 14-AUG-2000; 2000US-0225268.
XX
XX 14-AUG-2000; 2000US-0225270.
XX
XX 14-AUG-2000; 2000US-0225447.
XX
XX 14-AUG-2000; 2000US-0225757.
XX
XX 14-AUG-2000; 2000US-0225758.
XX
XX 14-AUG-2000; 2000US-0225759.
XX
XX 18-AUG-2000; 2000US-0226279.
XX
XX 22-AUG-2000; 2000US-0226681.
XX
XX 22-AUG-2000; 2000US-0226868.
XX
XX 22-AUG-2000; 2000US-0227182.
XX
XX 23-AUG-2000; 2000US-0227009.
XX
XX 30-AUG-2000; 2000US-0228924.
XX
XX 01-SEP-2000; 2000US-0229287.
XX
XX 01-SEP-2000; 2000US-0229343.
XX
XX 01-SEP-2000; 2000US-0229344.
XX
XX 01-SEP-2000; 2000US-0229345.
XX
XX 05-SEP-2000; 2000US-0229509.
XX
XX 05-SEP-2000; 2000US-0229513.
XX
XX 06-SEP-2000; 2000US-0230437.
XX
XX 06-SEP-2000; 2000US-0230438.
XX
XX 08-SEP-2000; 2000US-0231242.
XX
XX 08-SEP-2000; 2000US-0231243.
XX
XX 08-SEP-2000; 2000US-0231244.
XX
XX 08-SEP-2000; 2000US-0231413.
```

```
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0233066.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-465566/50.
DR P-PSDB; AAU23350.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
XX Claim 4; SEQ ID No 446; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AA540785-AA541684 represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 428 BP; 101 A; 94 C; 122 G; 106 T; 5 other;

alignment_scores:
Quality: 27.00 Length: 27
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAS41220 ..
Align seg 1/1 to: AAS41220 from: 1 to: 428

329 ProGluThrGluGlyIleLeuThrGluTyGlyValAspPheSerAspPh 345
|||||
43 CCGTAACACAGGAAGCAATACACAGAGTATGGCGTGATTCCTGATTT 92
|||||
345 eSerSerGluValLeuGluCysLeuProGln 355
|||||
93 CTCCTCAGAAAGTCTAGAAATGCTCTCTCAA 123

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23890
```

```

seq_documentation_block:
ID   AAZ23890 standard; cDNA; 1140 BP.
XX
AC   AAZ23890;
XX
DT   25-JAN-2000 (first entry)
XX
DE   Human LOBO cDNA fragment.
XX
KW   LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW   diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW   spondyloepiphyseal dysplasia; achondroplasia; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key      Location/Qualifiers
FT   CDS      1..784
FT           /*tag= a
FT           /product= "LOBO"
FT           /note= "partial sequence, no start codon given"
XX
XX   WO950284-A2.
XX
PD   07-OCT-1999.
XX
PF   26-MAR-1999; 99WO-EP02055.
XX
PR   27-MAR-1998; 98DE-1013799.
XX
PA   (ROSE/) ROSENTHAL A.
XX
PI   Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
WPI: 1999-601320/51.
XX
P-PSDB; AAY50353.
XX
XX   Nucleic acids encoding proteins which influence bone development,
XX   useful for treating and studying bone disorders -
XX
XX   Example 3; Page 66-67; 391pp; German.
XX
CC   This invention describes novel nucleic acids (I; designated LOBO (long
CC   bones)) encoding proteins influencing bone development in mammals. The
CC   proteins of the invention reduce and/or inactivate bone extension (i.e.
CC   development), with exception of the skull and have osteopathic activity.
CC   The nucleic acid molecules, proteins and antibodies can be used in
CC   diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC   and nucleic acid molecules, etc. are useful for production of transgenic
CC   animals, especially a transgenic mouse for the study of diseases
CC   associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC   achondroplasia. This sequence encodes the human LOBO protein described
CC   in the method of the invention.
XX
SQ   Sequence 1140 BP; 231 A; 355 C; 334 G; 220 T; 0 other;
XX

alignment_scores:
    Quality: 25.00      Length: 25
    Ratio: 1.000       Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23890 ..
Align seg 1/1 to: AAZ23890 from: 1 to: 1140

683 GlnPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSe 699
|||||
179 CAGTTCGGCAGCTACGGCTCATGTGCCCTGTACACACACTTCACCTC 228

699 rProileArgPheAlaAspVal 707
|||||
229 GCCCATCGCGCGCTTTCGGAGGTC 253

```

```

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23893
seq_documentation_block:
ID   AAZ23893 standard; DNA; 13330 BP.
XX
AC   AAZ23893;
XX
DT   25-JAN-2000 (first entry)
XX
DE   Human LOBO genomic DNA fragment.
XX
KW   LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW   diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW   spondyloepiphyseal dysplasia; achondroplasia; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO9950284-A2.
XX
PD   07-OCT-1999.
XX
PF   26-MAR-1999; 99WO-EP02055.
XX
PR   27-MAR-1998; 98DE-1013799.
XX
PA   (ROSE/) ROSENTHAL A.
XX
PI   Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
WPI: 1999-601320/51.
XX
XX   Nucleic acids encoding proteins which influence bone development,
XX   useful for treating and studying bone disorders -
XX
XX   Example 3; Page 118-125; 391pp; German.
XX
CC   This invention describes novel nucleic acids (I; designated LOBO (long
CC   bones)) encoding proteins influencing bone development in mammals. The
CC   proteins of the invention reduce and/or inactivate bone extension (i.e.
CC   development), with exception of the skull and have osteopathic activity.
CC   The nucleic acid molecules, proteins and antibodies can be used in
CC   diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC   and nucleic acid molecules, etc. are useful for production of transgenic
CC   animals, especially a transgenic mouse for the study of diseases
CC   associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC   achondroplasia. This sequence encodes the human LOBO protein described
CC   in the method of the invention.
XX
SQ   Sequence 13330 BP; 2787 A; 4068 C; 3977 G; 2498 T; 0 other;
XX

alignment_scores:
    Quality: 25.00      Length: 25
    Ratio: 1.000       Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23893 ..
Align seg 1/1 to: AAZ23893 from: 1 to: 13330

683 GlnPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSe 699
|||||
4013 CAGTTCGGCAGCTACGGCTCATGTGCCCTGTACACACACTTCACCTC 4062

699 rProileArgPheAlaAspVal 707
|||||
4063 GCCCATCGCGCGCTTTCGGAGGTC 4087

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23900
seq_documentation_block:

```

```
ID AAZ23900 standard; DNA; 49999 BP.
XX
AC AAZ23900;
XX
DT 25-JAN-2000 (first entry)
DE Human LOBO homologue genomic DNA fragment 2.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
DE diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
XX spondyloepiphyseal dysplasia; achondroplasia; human; ds.
XX
OS Homo sapiens.
XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX
PS Example 3; Page 244-271; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes a human LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 11432 A; 13017 C; 12505 G; 13045 T; 0 other;

alignment_scores:
    Quality: 25.00      Length: 25
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23900 ..
Align seg 1/1 to: AAZ23900 from: 1 to: 49999

683 GlnPheArgHisTyrAlaLeuasnValProLeuTyrThrHisPheThrSe 699
|||||
48528 CAGTTCGGCGCACTACGGCTCAATGTCCTGACACACACTTCACCTC 48577

699 rProtleArgArgPheAlaaspVal 707
|||||
48578 GCCCATCGCGCGCTTTCGCCGACGTC 48602

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ23899

seq_documentation_block:
ID AAZ23899 standard; DNA; 49999 BP.
XX
AC AAZ23899;
XX
DT 06-OCT-2000 (first entry)
DE Human secreted protein 5' EST, SEQ ID NO: 34871.
XX
XX
```

```
DT 25-JAN-2000 (first entry)
XX
DE Human LOBO homologue genomic DNA.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
DE diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
XX spondyloepiphyseal dysplasia; achondroplasia; human; ds.
XX
OS Homo sapiens.
XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX
PS Example 3; Page 215-243; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes a human LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13475 A; 10783 C; 11425 G; 14316 T; 0 other;

alignment_scores:
    Quality: 19.00      Length: 19
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-647-377-9 x AAZ23899 ..
Align seg 1/1 to: AAZ23899 from: 1 to: 49999

438 ValValProMetLeuProArgLeuLeuCysgluLeuCysSerLeuAs 454
|||||
13422 GTGGTCCCATGCTTCCAGGCTGCTGTGTGAGAGAGCTGTGCAGCTCAA 13471

454 nProMet 456
|||||
13472 CCCCATG 13478

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ30796

seq_documentation_block:
ID AAC30796 standard; cDNA; 331 BP.
XX
AC AAC30796;
XX
DT 06-OCT-2000 (first entry)
DE Human secreted protein 5' EST, SEQ ID NO: 34871.
XX
XX
```

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
OS  
XX  
PN EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 34871; 71pp + CD-ROM; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
SQ Sequence 331 BP; 73 A; 112 C; 84 G; 59 T; 3 other;

alignment\_scores:  
Quality: 18.00 Length: 18  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-647-377-9 x AAC30796 ..

Align seg 1/1 to: AAC30796 from: 1 to: 331

645 ThrPheGlyAspAspTyrSerLeuAlaArgGlyGluValLeuThrAs 661  
|||||  
240 ACATTTGGAGTACAGTACTACTGCGCCGACAGGAGTCTCACCAA 289

661 nMet 662  
|||||  
290 CATG 293

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK60530

seq\_documentation\_block:

ID AAK60530 standard; cDNA; 925 BP.

XX

AC AAK60530;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:5590.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ss.  
XX Homo sapiens.  
OS  
XX  
PN WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI: 2001-483426/52.  
DR P-PSDB; AAK87749.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX Claim 1; SEQ ID NO 5590; 3071pp + Sequence Listing; English.  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX SQ Sequence 925 BP; 216 A; 258 C; 262 G; 185 T; 4 other;

alignment\_scores:  
Quality: 16.00 Length: 16  
Ratio: 1.000 Gaps: 0  
Percent similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-647-377-9 x AAK60530 ..  
Align seg 1/1 to: AAK60530 from: 1 to: 925  
805 pheGlnLysValGlyLysLysProGluLeuThrLeuValTrpGluPro 820  
|||  
82 TTCCAGAGGTNGGCAAGAACCGGAACACACGTCGTGGAGCCT 129





OM of: US-09-647-377-9 to: N\_Geneseq\_032802.\* out\_format : pfs  
 Date: Jul 8, 2002 12:57 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL=framer\_p2n.model -DEV=xlh  
 -Q/cgn2.1/USPTC\_spool/US09647377/runat\_08072002\_075945\_22334/app\_query.fasta\_1.940  
 -DB=N\_Geneseq\_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
 -USER=US09647377\_RCGN1\_1.498 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
 Query: US-09-647-377-9  
 Query length: 870  
 Database: N\_Geneseq\_032802.\*  
 Database sequences: 1736436  
 Database length: 858457221  
 Search time (sec): 240.820000

```
score_list:
Sequence      Strd Orig      ZScore      EScore Len      Documentation
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23894 + 4521.00 7779.80 0.0
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23898 + 2185.50 3747.59 2.1e-20
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23889 + 1944.50 3332.48 2.8e-17
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23890 + 1083.00 1846.32 1.7e-94
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL11059 + 994.00 1680.70 2.9e-8
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL02002 + 972.50 1637.19 7.6e-8
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL11058 + 920.50 1547.33 7.7e-7
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL02003 + 859.50 1448.75 2.4e-7
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAZ23891 + 816.50 1024.15 1.1e-4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:ABA03041 + 595.00 921.00 2.1e-42
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23892 + 571.00 910.47 6.4e-43
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23897 + 571.00 919.86 6.9e-43
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH53049 + 568.00 947.19 2.1e-4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH54223 + 567.00 942.34 3.8e-4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAH54223 + 566.50 933.82 1.1e-43
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:AAV74495 + 541.00 889.44 3.4e-41
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV52241 + 532.00 866.80 6.2e-40
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAAO5471 + 521.50 866.91 6.1e-40
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAAC81914 + 503.50 780.45 4.0e-35
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23893 + 480.00 774.96 8.1e-35
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23900 + 480.00 759.58 5.8e-34
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAAC30796 + 431.00 733.24 1.7e-32
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ41220 + 424.50 719.01 1.1e-3
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23891 + 424.50 663.60 1.3e-28
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23896 + 424.50 663.60 1.3e-28
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:AAI67439 + 419.50 690.34 4.2e-30
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ79601 + 419.00 680.46 1.5e-29
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ54518 + 418.50 688.69 5.2e-30
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF28548 + 416.00 641.30 2.3e-2
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAF21609 + 412.50 620.20 3.4e-26
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAF21610 + 412.50 620.20 3.4e-26
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ54516 + 407.50 669.65 5.9e-29
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH81459 + 406.50 631.03 8.4e-27
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ54517 + 403.50 665.19 1.1e-28
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23899 + 383.00 591.84 1.3e-24
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:AAZ58840 + 383.00 563.31 5.0e-23
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAH81490 + 373.00 535.46 1.8e-21
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH86295 + 370.50 594.54 9.1e-2
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAFL4474 + 367.00 613.82 7.6e-26
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23903 + 366.00 562.44 5.6e-23
```

```
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ42031 + 331.50 538.97 1.1e
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ42032 + 331.50 538.97 1.1e
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ42033 + 328.50 544.04 5.9e
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ42034 + 309.00 482.08 1.7e
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ42035 + 277.00 453.99 6.1e
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23894
```

```
seq_documentation_block:
ID AAZ23894 standard; cDNA to mRNA; 3100 BP.
XX
AC AAZ23894;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO homologue cDNA to mRNA.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ss.
XX
OS Mus musculus.
```

```
XX Key Location/Qualifiers
XX CDS 125..2737
XX FT /*tag= a
XX FT /product= "LOBO homologue"
XX
XX PN W09950284-A2.
XX
XX PD 07-OCT-1999.
XX
XX PF 26-MAR-1999; 99WO-EP02055.
XX
XX PR 27-MAR-1998; 98DE-1013799.
```

```
XX (ROSE/) ROSENTHAL A.
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI: 1999-601320/51.
XX P-PSDB; AAY50354.
XX
XX Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX Claim 1b; Page 126-130; 391pp; German.
```

```
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes a murine LOBO protein described
XX in the method of the invention.
```

XX Sequence 3100 BP; 783 A; 795 C; 812 G; 710 T; 0 other;

```
alignment_scores:
Quality: 4521.00 Length: 870
Ratio: 5.197 Gaps: 0
```

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-647-377-9 x AAZ23894 ..

Align seg 1/1 to: AAZ23894 from: 1 to: 3100

1 MetAsnHisProAspTyrIleLysLeuAsnLeuArgSerProGlyThrProAr 17  
|||||  
125 ATGAACCATCTGACTACAGCTGAACTTCGGTCTCCGGGGACCCCHG 174  
17 gGlyValSerSerValIleGlyProSerAlaValGlyAlaSerProGlyA 34  
|||||  
175 AGGTGTCTCTGTGGTGGCCGAGTGTCTGTGGTGTCTGCCAGGTG 224  
34 sPLysLysSerLysAsnLysSerMetArgGlyLysLysLysSerIlePhe 50  
|||||  
225 ACAAAAGTCAAGAACAGTCCATGCGGGGGAAGAAAGACATATTT 274  
51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67  
|||||  
275 GAAACCTACATGTCCAAGGAGGATGTTTCAGAAGGCTTGAAGAGGAAC 324  
67 rLeuIleGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84  
|||||  
325 ACTTATCCAGGGTGATTAGAGATCAACCCAAAGAGTTTCATGAAGCT 374  
84 heIleProSerProAspGlyAspArgAspIlePheIleAspGlyValVal 100  
|||||  
375 TCATTCCTCTCCGGATGGTGTGCGGACATTTTATTATGTATGGAGTTGT 424  
101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuPr 117  
|||||  
425 GCTCGTAATAGAGCCTTAAATGGGACCTTGTGGTGTAAACTGCTTCC 474  
117 oGluAspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIleG 134  
|||||  
475 TGAGGATCAGTGGAGGAGGATTAACCCAGAGAGCAATGCAAGAAATAG 524  
134 luAlaThrTyrGluAlaAspIleProGluGlyCysGlyHisHisPro 150  
|||||  
525 AAGCTACTTATGAAGCTGACATCCCTGAAGAGGGCTGTGGACATCACCC 574  
151 LeuGlnGlnSerArgLysGlyTrpSerGlyProAspValIleIleGluAl 167  
|||||  
575 CTGACAGTCCCGGAAAGGTGGAGTGGTCCCTGATGTCATTATAGAGGC 624  
167 aGlnPheAspSerAspSerGluAspArgHisGlyAsnThrSerGlyL 184  
|||||  
625 TCAGTTTGATGACAGCGACTCAGAAATGACATGGCAACACACAGTGCC 674  
184 euValAspGlyValLysLysLeuSerIleSerThrProAspArgGlyLys 200  
|||||  
675 TGGTTGATGGTGTAGAAATGTCAATCTCTACTCCTGACAGAGGAAAA 724  
201 GluAspSerSerThrProValMetLysAspGluAsnThrProIleProGl 217  
|||||  
725 GAAGATTCTAGTACTCCAGTTATGAAAGATGAGAACACCCCATACACA 774  
217 nAspThrArgGlyLeuSerGluLysSerLeuGlnLysSerAlaLysValV 234  
|||||  
775 GGACACAAAGAGCTTATCAGAAAGTCACTTCAGAAATCAGCAAAAGGTG 824  
234 alTyrIleLeuGluLysLysHisSerArgAlaAlaThrGlyIleLeuLys 250  
|||||  
825 TTTTACATCTTGGAGAAAAGCATCTCGAGACGAACTGGCATCCTGAAA 874  
251 LeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaLeuPheSe 267  
|||||  
875 CTCTTGGCTGATAGAACAGTGACCTGTTTAAAGAAATACGCCCTGTTTC 924  
267 rProSerAspHisArgValProArgIleTyrValProLeuLysAspCysp 284  
|||||  
925 TCCTTCAGACCAACCCAGGTACCTAGAAATTTACGTACCTCTCAAGGACTGTC 974  
284 roGlnAspPheMetThrArgProLysAspPheAlaAsnThrLeuPheIle 300  
|||||  
975 CCAGAGACTTCATGACCCGACCTTAAGACTTTTGCACACAGCTGTTCATC 1024  
301 CysArgIleIleAspTrpLysGluAspCysAsnPheAlaLeuGlyGlnLe 317  
|||||

1025 TGCCGCATCATAGACTGGAAGGAGGACTGTAATTTTGCCCTGGGGCACT 1074  
317 uAlaLysSerLeuGlyGlnAlaGlyGluIleGluProGluThrGluGlyI 334  
|||||  
1075 GGTAAAGAGTCTTGGCAGGCTGGTGAATCGAGCCTGAACACAGAGGA 1124  
334 leLeuThrCluTyrGlyValAspPheSerAspPheSerSerGluValLeu 350  
|||||  
1125 TACTGACAGAATATGGGTGGACTTCTCTGATTTCTCTCAGAAAGTCTTT 1174  
351 GluCysLeuProGlnSerLeuProTrpThrIleProProAspGluValGl 367  
|||||  
1175 GAATGTCTCCCTCAAGCCTGCCCTGGACAA TCCACACCTGATAGGTGG 1224  
367 yLysArgArgAspLeuArgLysAspCysIlePheThrIleAspProSerT 384  
|||||  
1225 CAAGAGAAGACCTTAAGGAAGACTGTATCTTCACCATTTGATCCATCAA 1274  
384 hrAlaArgAspLeuAspAlaLeuAlaCysArgArgLeuThrAspGly 400  
|||||  
1275 CTGCTCGGACCTTGTATGATGCTCGCTCGCTGAGGCGGCTCACTGATGGC 1324  
401 ThrPheGluValGlyValHisIleAlaAspValSerTyrPheValProGl 417  
|||||  
1325 ACCTTCGAAGTGGCGTCCACATCGCCGATGTGATTTCTTCTCTGA 1374  
417 uGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrL 434  
|||||  
1375 GGGATCTCTTTGGATAAGTAGTCTGCTGAGAGAGCCACAAGTGTCTACT 1424  
434 euValGlnLysValValProMetLeuProArgLeuLeuCysGluLeu 450  
|||||  
1425 TGSTCCAGAAGTGGTCCCATGCTTCCAGGCTTCTGTGTGAGGAACTC 1474  
451 CysSerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpIly 467  
|||||  
1475 TGCAGCCTCAACCCCATGACTACAAAGCTGACCTTCTCTGTGATCTGAA 1524  
467 sleuThrProGluGlyLysIleLeuGluGluTrpPheGlyArgThrIleI 484  
|||||  
1525 GCTGACCCCTGAAGGCAAGATCCTTGAAGAGTGGTGTGGCGCACATATCA 1574  
484 leArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGlu 500  
|||||  
1575 TCCGTCTTCACCAAACTAGCTACGACCATGCCAGAGCATGATCGAA 1624  
501 AsnProThrGluLysIleProGluGluGluLeuProProIleSerProGl 517  
|||||  
1625 AATCCAACTGAGAAGATCCCTGAGGAAGAGCTTCCCCCAATTTCTCCGA 1674  
517 uHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleA 534  
|||||  
1675 GCACAGCGTCGAGGAGGTGCACGAGCAGTCTGAACCTGCACAGCATTTG 1724  
534 laLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAsp 550  
|||||  
1725 CAAGCAACTCCCGCCGACGCTTTGTAGATGGCGCACTCCGTTTAGAT 1774  
551 GlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnGl 567  
|||||  
1775 CAGCTGAAGCTTGTCTTTTACTCTGGACCATGAGACTGGACTGCTCAAGG 1824  
567 yCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPheM 584  
|||||  
1825 ATGTCACATCTATAGTACCGAGACAGCAACAAGCTTGTAGAGGAGTTCA 1874  
584 etLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhePro 600  
|||||  
1875 TGCTCTCTGGGCAACATGGCGGTGGCCCAAGATCTTCCGCACTTCCCT 1924  
601 GluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSe 617  
|||||

1925 GAGAGGCCCTGCTGCCGCGCATCCCCACACAGACGAAGATGCTCAG 1974  
617 rAspLeuValGluPheCysaspGlnMetGlyLeuProMetaspValSerS 634  
1975 TGACCTGTGTGAGTTCGTGACACAGATGGGGTCCCATGTGATGTCAGT 2024  
634 eRAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspLys 650  
2025 CTGCAGGGGCCCTAAATAAAGCCTGACTAAGACATTGGAGATGACAAG 2074  
651 TyrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMe 667  
2075 TACTCTCTGGCCCGAGAGAGGTGCTCACCACATGTACTCCCGGCCAT 2124  
667 tGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluGlnP 684  
2125 GCAGATGGCACTGTACTTCTGCTCTGGATGCTGCAGGACGAGCAGT 2174  
684 heaRgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerPro 700  
2175 TCCGGCATATGCTCTCAACGTTCCCTCTACACACACTTCCCTCTCCC 2224  
701 IleArgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLe 717  
2225 ATCCGCGCTTTGCTGAGTCATAGTCACCGCTCTCTGGCTGCTCT 2274  
717 uGlyTyrSerGluGlnProaspValGluProaspThrLeuGlnLysGlnA 734  
2275 GGGTACAGTGAACAGCCAGATGTGGAGCTGATACCTACAGAGCAAG 2324  
734 laAspHisCysAsnAspArgMetAlaSerLysArgValGlnGlnGluLeu 750  
2325 CTGACCACTGCAATGACCGTCGATGCTTCCAAACAGTGTGACGAGCTC 2374  
751 SerIleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuG 767  
2375 AGCATCGGCTCTTCTTTCGAGTCTTAGTAAAGGAGAGTGGCCCCCTGGA 2424  
767 uSerGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuV 784  
2425 GTCCGAAGCCATGTGTGGTGTCTTCTGAACCAAGCTTCGACGTGCTGG 2474  
784 alLeuArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeu 800  
2475 TGTGCGCTTTGGGTGCAGAGCGCATCTACTGCAATGCACCTGCCCTG 2524  
801 ArgSerTyrSerPheGlnLysValGlyLysLysProGluLeuThrLeuVa 817  
2525 CGATCTACAGCTTCCAGAGGTGGGAAGAGCCAGAGCTCACTCTGT 2574  
817 lTrpGluProaspAspLeuGluGluProThrGlnGlnValIleThrI 834  
2575 TTGGGAGCTGATGACCTTGAAGAGGAGCAACACAGAGGTCAATCACA 2624  
834 lePheSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLys 850  
2625 TCTTCAGCTGTGTGATGCTGCTGCGAGCAGAGGCCACAGCCCTCAAG 2674  
851 TyrSerAlaIleLeuLysArgProGlyLeuGluLysAlaSerAspGluCl 867  
2675 TACAGTGTATCTCCTGAACGACAGCGCTGGAGAAGGCGCTGTGTAGGA 2724  
867 uProGluAsp 870  
2725 GCCTGAGGAC 2734

seq\_name: /SIDS1/c9gdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23898

seq\_documentation\_block:

ID AAZ23898 standard; cDNA to mRNA; 1784 BP.

XX AC AAZ23898;

XX

DT 25-JAN-2000 (first entry)  
XX Human LOBO homologue cDNA to mRNA.  
DE LOBO; long bones; bone development; bone extension; skull; osteopathic;  
XX diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
KW spondyloepiphyseal dysplasia; achondroplasia; human; ss.  
KW Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FT CDS L..1454  
FT /\*tag= a  
FT /product= "LOBO homologue"  
FT /note= "Partial sequence, no start codon given"  
XX  
XX WO950284-A2.  
XX  
XX 07-OCT-1999.  
XX  
XX 26-MAR-1999; 99WO-EP02055.  
XX  
XX 27-MAR-1998; 98DE-1013799.  
XX (ROSE/) ROSENTHAL A.  
XX  
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;  
XX  
XX WPI; 1999-601320/51.  
XX p-PSDB; AAY50355.  
XX  
XX Nucleic acids encoding proteins which influence bone development,  
XX useful for treating and studying bone disorders -  
XX  
XX Claim 1b; Page 211-213; 391pp; German.  
XX  
XX This invention describes novel nucleic acids (1; designated LOBO (long  
XX bones)) encoding proteins influencing bone development in mammals. The  
XX proteins of the invention reduce and/or inactivate bone extension (i.e.  
XX development), with exception of the skull and have osteopathic activity.  
XX The nucleic acid molecules, proteins and antibodies can be used in  
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
XX and nucleic acid molecules, etc. are useful for production of transgenic  
XX animals, especially a transgenic mouse for the study of diseases  
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and  
XX achondroplasia. This sequence encodes a human LOBO protein described  
XX in the method of the invention.  
XX  
XX Sequence 1784 BP; 385 A; 537 C; 504 G; 358 T; 0 other;

alignment\_scores:  
Quality: 2185.50 Length: 478  
Ratio: 4.825 Gaps: 1  
Percent Similarity: 94.770 Percent Identity: 88.494  
alignment\_block:  
US-09-647-377-9 x AAZ23898 ..  
Align seg 1/1 to: AAZ23898 from: 1 to: 1784  
402 PheGluValGlyValHisIleAlaAspValSerTyrPheValProGluG 418  
6 TTCAAAGTGGGAGTTCACATTGCTGACGTGAGTTACTTTGTCGGAGGG 55  
418 ySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrLeuV 435  
56 ATCTGATCTGGATAAAGTGGCTGCCGAGAGGGCTACAAGCGTCTACTTGG 105  
435 alGlnLysValValProMetLeuProArgLeuLeuLysGluGluLeuCys 451  
106 TTCAAAGAGGTGGTCCCATGCTTCCAGCGTGTGTGTGAGGAGCTGTGC 155

```
452 SerLeuAsnProMetThrAspLysLeuThrPheSerValIleThrLysLe 468
|||||
156 AGCCTCAACCCATGTCGCGAAGCTGACCTTCTGTGTGATCTGGACACT 205
|||||
468 uThrProGluGlyLysIleLeuGluThrPheGlyArgThrIleLea 485
|||||
206 GACTCCAGAGGGCAAGATCCTTGAATGGTGTGGCGGACCATCATCC 255
|||||
485 tGSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGluAsn 501
|||||
256 GCTCTCGACCAACTTAGCTAGGAGTGCACAGACATGATGAAGC 305
|||||
502 ProThrGluLysIleProGluGluLeuProProIleSerProGluHi 518
|||||
306 CCAACTGAGAAATCCCTGCGAAGAGGCTGCCGCCCATTTCCACAGCA 355
|||||
518 sSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleAla 535
|||||
356 TAGCAGCGAGGAGGTACACGCGCGTCTTGAATCTCCAGGAATTGCCA 405
|||||
535 yGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAspGln 551
|||||
406 AGCAGTTACGCCACGACGCGCTTGTGGACGGCGCACTTCGTTGGATCAG 455
|||||
552 LeuLysLeuAlaPheThrLeuAspHisGlnThrGlyLeuProGlnGlyCy 568
|||||
456 CTAAAGCTTGCTTCTACTGTGACCACGAGACCGGATTCCTCAAGATG 505
|||||
568 sHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPheMetL 585
|||||
506 TCATATCTATGAGTACGCGGAGAGCAACAAGCTCGTGAGGAGTTTCATGC 555
|||||
585 euLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPheProGlu 601
|||||
556 TCTTGGCCACATGGCAGTGGCCACAAAGATCCACCGCGCTTCCCGGAG 605
|||||
602 GlnAlaLeuLeuArgHisProProGlnThrLysMetLeuSerAs 618
|||||
606 CAGGCGCTGTGCGCGGACCCCGCCGCCCCAACAAAGATGCTCAGTGA 655
|||||
618 pLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSerA 635
|||||
656 CTGCTGGAATTCGCGACACAGATGGGCTGCGCGTGGACTTCAGTCCG 705
|||||
635 laGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspLysTyr 651
|||||
706 CAGGAGCCCTCAATAAAGCCTGACCCCAACATTGTGAGATGACAAAGTAC 755
|||||
652 SerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMetGl 668
|||||
756 TCACTGCGCCGCAAGGAGGTGCTACCAACATGTGCTCCGCGCCATGCA 805
|||||
668 nMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluInPheA 685
|||||
806 GATGGCCTTGCTGCTGCGGCTGCTGACGAGCCAGCGCATGTC 855
|||||
685 rGHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIle 701
|||||
856 GGCACCTACGGGCTCAATGTGCCCCGTGTACACACATTCACCTGCGCCATC 905
|||||
702 ArgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGl 718
|||||
906 GCGCGCTTGCGAGCTCTGCTGACGCGCTCTGCTGCTGCTGCTGCTGCTG 955
|||||
718 yTyrSerGluGlnProAspValGluProAspThrLeuGlnLysGlnAlaA 735
|||||
956 CTATAGGAGCGACTAGACATGGCGCGGATACCTTCAGAACACAGCGCG 1005
|||||
735 spHisCysAsnAspArgMetAlaSerLysArgValGlnGluLeuSer 751
|||||
1006 ACCACTAACGACCGCGCTGCGTCCAAAGCGCGTGCAGGAGTCACT 1055
|||||
752 IleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuGluSe 768
```

```
|||||
1056 ACCAGTCTCTTTGCTGTGTTCTGTTCAAGGAGAGTGCGCCCTGGAGTC 1105
|||||
768 rGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeuValL 785
|||||
1106 AGAAGCCATGGTGATGGGCATCCTGGAAGCAAGCTTCGACGCTGCTGGTGC 1155
|||||
785 euArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArg 801
|||||
1156 TGGCTACCGCTGCGAGAGCGCATCTACTGCAACGCACATGGCCCTCGG 1205
|||||
802 SerTyrSerPheGlnLysValGlyLysProGluLeuThrLeuValTr 818
|||||
1206 TCCACCACTTCCAGAAGGTGGCAAGAGCGGAACCTCACGCTGGTCTG 1255
|||||
818 pGluProAspAspLeuGluGluProThrGlnGlnValIleThrIleP 835
|||||
1256 GGAGCCTGAGGACATGGAGCAGGAGCGACGACGAGGTCTCATCCATCT 1305
|||||
835 heSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuLysTyr 851
|||||
1306 TCAGCCTGTGTGAGGTGCTCTGACGACGAGAGTCCACAGCCCTCAAGTAC 1355
|||||
852 SerAlaIleLeuLysArgProGlyLeu..... 860
|||||
1356 AGCGCATCTGAGCGCGCCAGGCCACCCAGGGCCACCTGGGCGCTGAGAA 1405
|||||
861 ...GluLysAlaSerAspGluGluProGluAsp 870
|||||
1406 GGAGGAGGAGGAGTCTGACGCTGAGCGCCGAGGAC 1439
|||||
```

seq\_name: /SIDS1/gcdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23889

seq\_documentation\_block:

ID AAZ23889 standard; cDNA; 1550 BP.

XX AAZ23889;

XX AC AAZ23889;

XX DT 25-JAN-2000 (first entry)

XX XX Murine LOBO cDNA fragment.

XX DE LOBO; long bones; bone development; bone extension; skull; osteopathic;

XX KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;

XX KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

XX FT CDS l..1183

XX FT /\*tag= a

XX FT /product= "LOBO"

XX FT

XX PN WO9950284-A2.

XX XX 07-OCT-1999.

XX PD 26-MAR-1999; 99WO-EP02055.

XX PF 27-MAR-1998; 98DE-1013799.

XX PR (ROSE/) ROSENTHAL A.

XX PA Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;

XX PI WPI; 1999-601320/51.

XX DR P-PSDB; AAY50352.

XX XX Nucleic acids encoding proteins which influence bone development,

XX PT useful for treating and studying bone disorders -

XX XX Example 3; Page 62-64; 391pp; German.

XX PS

XX XX

This invention describes novel nucleic acids (I; designated LOBO (long bones)) encoding proteins influencing bone development in mammals. The proteins of the invention reduce and/or inactivate bone extension (i.e. development), with exception of the skull and have osteoplastic activity. The nucleic acid molecules, proteins and antibodies can be used in diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods and nucleic acid molecules, etc. are useful for production of transgenic animals, especially a transgenic mouse for the study of diseases associated with bone development, e.g. spondyloepiphyseal dysplasia and achondroplasia. This sequence encodes the murine LOBO protein described in the method of the invention.

Sequence 1550 BP; 352 A; 433 C; 412 G; 353 T; 0 other; xx

alignment_scores:			
Quality:	1944.50	Length:	448
Ratio:	4.999	Gaps:	2
percent Similarity:	86.830	Percent Identity:	86.830

alignment\_block:

US-09-647-377-9 x AAZ23889

Align seq 1/1 to: AAZ23889 from: 1 to: 1550

423 LysValAlaAlaGluArgAlaThrSerValTyrLeuValGlnLysValVa 439  
14 AAAGTAGCTGCTGAGAGAGCCAAAGTGCTACTTTGGTCCAAAGTGGT 63  
439 lPrometLeuProArgLeuLeuCysGluLeuCysSerLeuAsnProM 456  
64 CCCCATGCTCCACGGCTTGTGTGAGGAACCTGTCAGGCTCAACCCCA 113  
456 etThrAspLysLeuThrPheSerValIleTrrLysLeuThrProGluGly 472  
114 TGACTGACAGCTGACCTTCTGTGTGATCTGAAGCTGACCCCTGAAGGC 163  
473 LysIleLeuGluLutrpPheGlyArgThrIleIleArgSerCysThrLy 489  
164 AAGATCTTGAAGAGTGGTTTGGCCGCACATATCATCCGCTCTTGACCAA 213  
489 sLeuSerTyrAspHisAlaGlnSerMetIleGluAsnProThrGluLysI 506  
214 ACTGAGCTACGACCATGCGCCAGCAGCATGATCGAANAATCCAACTGAGAAGA 263  
506 leProGluGluLeuLeuProPilesSerProGluHisSerValGluGlu 522  
264 TCCCTGAGGAAGAGCTTCCCCCAATTTCTCCAGACACAGCGTCGAGGAG 313  
523 ValHisGlnAlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgAr 539  
314 GTGCACACAGCAGCTCTCAACCTGCACAGCATTTGCAAGAACTCCGCGC 363  
539 gGlnArgPheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaP 556  
364 CCACGCTTTGTAGATGGCGACATCCGCTTTAGATCAG..... 400  
556 heThrLeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrGlu 572  
400 ..... 400  
573 TyrArgAspSerAsnLysLeuValGluLupheMetLeuLeuAlaAsnMe 589  
401 .....GAGTTCATGCTCTGGCCACAT 423  
589 talValAlaHisLysIlePheArgThrPheProGluGlnAlaLeuLeuA 606  
424 GCGCGTGGGCCCAAGATCTTCCGACCTTCCCTGAGCAGCCCTGTGTC 473  
606 rGArgHisProProGlnThrLysMetLeuSerAspLeuValGluPhe 622  
474 SCCGGCATCCCCACCACAGACGAGAGATGCTCAGTCACCTGCTGGAGTTC 533

623	CysAspGlnMetGlyLeuProMetAspValSerSerAlaGlyAlaLeuAs	639
524	TGTACCCAGATGGGGTGGCCATGGATGTCTCAGCTCTGCAGGGGCCCTAAA	573
639	nLysSerLeuThrLysThrPheGlyAspAspLysTyrSerLeuAlaArgL	656
574	T.....	574
656	ysGluValLeuThrAsnMetTyrSerArgProMetGlnMetAlaLeuTyr	672
575	.....ATGGCACTGTAC	586
673	PheCysSerGlyMetLeuGlnAspGlnGluInPheArgHisTyrAlaLe	689
587	TTCTGCTCTGGATGTCGAGGACCAGGAGCTTCGGCATTATGCTCT	636
689	uAsnValProLeuTyrThrHisPheThrSerProIleArgArgPheAlaA	706
637	CAACGTTCCCTCTACACACATTCACCTCCCATCCGCCGCTTGGTG	686
706	spValIleValHisArgLeuLeuAlaAlaLeuGlyTyrSerGluIn	722
687	ACGTATAGTGCACCGCCTCTGGTGTCTCTGGCTACAGTGAACAG	736
723	ProAspValGluProAspThrLeuGlnLysAlaAspHisCysAsnAs	739
737	CCAGATGTGGAGCCTGATACCTACAGAGCAGCTGACCTGGCATGA	786
739	pArgArgMetAlaSerLysArgValGlnGluLeuSerIleGlyLeuPheP	756
787	CGTCCGATGGCTCCAAACGTGTCCAGGAGCTCAGCATCGGCTCTTCT	836
756	heAlaValLeuValLysGluSerGlyProLeuGluSerGluAlaMetVal	772
837	TGCGATTCTATTAAGGAGAGTGGCCCCCTGGAGTCCGAGCCATGGT	886
773	MetGlyValLeuAsnGlnAlaPheAspValLeuValLeuArgPheGlyVa	789
887	ATGGGTGTCTTGAAACCAAGCTTCCAGCTGTGTGTCTGGCTTTGGGGT	936
789	IGlnLysArgIleTyrCysAsnAlaLeuAlaLeuArgSerTyrSerPheG	806
937	GCAGAGAGCGCATCTACTGCATATGCATCGGCCCTCGCATCTACAGCTTC	986
806	InLysValGlyLysLysProGluLeuThrLeuValTrpGluProAspasp	822
987	AGAAAGTGGGGAAGAGCCAGAGCTCCTGTGTGTGGAGAGCTCATGAC	1036
823	LeuGluGluGluProThrGlnGlnValIleThrIlePheSerLeuValAs	839
1037	CTTGAGAGGAGCCACACAGCAGGTCATCACCATTCTCAGCGCTGGTGA	1086
839	pValValLeuGlnAlaGluAlaThrAlaLeuLysTyrSerAlaIleLeuL	856
1087	TGTGTCTCTGAGCGCAGAGGCCACAGCCCTCAAGTACAGTGTCTTCCTGA	1136
856	ysArgProGlyLeuGluLysAlaSerAspGluGluProGluAsp	870
1137	AGCGCAGCGCTCGAAGAGGCTGTGATGAGGAGCCTGAGGAC	1180

seq name: /STDS1/acadata/hold-geneseg/geneseg-emb1/NA1999 DAT:AAZ23890

seq documentation block:

seq\_documentation\_block:  
ID AAZ23890 standard; cDNA: 1140 BP.

[illegible]

AAZ23890;

XX  
XX  
XXXXXXX

DT 25-JAN-2000 (first entry)

XXXXXX

Human LOBO cDNA fragment.

[illegible]

**KW** LOBO; long bones; bone development; bone extension; skull; osteopathic; osteopathy

diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84



XX 23-MAR-2001; 2001WO-US09231.  
 XX PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX PA  
 XX (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR P-PSDB; ABB66956.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 27659; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3119 BP; 803 A; 805 C; 839 G; 672 T; 0 other;

## alignment\_scores:

Quality: 994.00 Length: 861  
 Ratio: 1.912 Gaps: 23  
 Percent Similarity: 60.395 Percent Identity: 31.010

## alignment\_block:

US-09-647-377-9 x ABL11059 ..

Align seg 1/1 to: ABL11059 from: 1 to: 3119

20 SerSerValValGlyProSerAlaValGlyAlaSerProGlyAsp.LysL 36  
 ||||| ||| : : : ||| ||| : : |||  
 660 AGCAGCGGAGTACGTGAATCTCTCGAGGATTTCCACTTCTGTGGATA 709  
 36 ysSerLysAsnLysSerMetArgGlyLysLysLysSerIle.....Phe 50  
 || : : : ||| :  
 710 AGCTGTGCGCACAAACATTCGACAGCGAGAAGATGGTCTGCCCCAGTAC 759  
 51 GluThrTyrMetSerLysGluAspValSerGluGlyLeuLysArgGlyTh 67  
 :  
 760 CCGCCGCATCTCAGCATGAAGAGCTCTGGAAGGAGCTCGCCCAAAATAA 809  
 67 rLeuLeGlnGlyValLeuArgIleAsnProLysLysPheHisGluAlap 84  
 ||||| :  
 810 GCTACTACAGGCACCTTTCAGGCATCCAGGAAACTACTTGGAGGCA 859  
 84 heLleProSerProAspGlyAspArgAspIlePheLeaspGlyValVal 100  
 :  
 860 CCGTCAATGTAGAGAATTCGAAGAGGGATCTCTTATCCAAGGACGCGAG 909  
 101 AlaArgAsnArgAlaLeuAsnGlyAspLeuValValLysLeuLeuPr 117  
 :  
 910 TCTCTTANCGCGCGTGTGGACCTGGTGGCCGTGGAGCTTTTGGC 959  
 117 oGluAspGlnTrpLysAlaValLysProGluSerAsnAspLysGluIleG 134  
 |||| : : : ||| :  
 960 CGAGCGGAGTGTGTACGA.....CCAGTGAATGTGTACTGGAGAGA 1003  
 134 luAlaThrTyrGluAlaAspIleProGluGlyCysGlyHisHisPro 150

: : : : ||| : : : : |||  
 1004 AGAATGTCTATGCCGATGAAGTGCC ..... 1029  
 151 LeuGlnGlnSerArgLysGlyTrpSerGlyProAspValIleIleGluAl 167  
 1029 ..... 1029  
 167 aGlnPheAspSerAspSerGluAspArgHisGlyAsnThrSerGlyL 184  
 ||||| :  
 1030 ..... AGCGAGGAAGGCCCAAGGATGAAATGAGA 1060  
 184 euValAspGlyValLysLysLeuSerIleSerThrProAspArgGlyLys 200  
 :  
 1061 TGTGTAATCAGTGGCTGCCGTGCTTATCCGA ..... 1095  
 201 GluAspSerSerThrProValMetLysAspGluAsnThrProIleProG 217  
 ||||| :  
 1096 ..... GAGCGCACGCC ..... 1107  
 217 nAspThrArgGlyLeuSerGluLysSerLeuGlnLysSerAlaLysValV 234  
 ..... ACAGGCGGAATTG 1120  
 234 alTyrIleLeuGluLysLysHisSerArgAlaAlaThrGlyIleLeuLys 250  
 || :  
 1121 TCGGTATTGTGCGACGAAATGG...CGACAGTATTGCGGCATCTCTGCAG 1167  
 251 .....LeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaLe 265  
 ||||| :  
 1168 CTAAGCTTATAGAGATACAAATCGGCAC .....AT 1199  
 265 uPheSerProSerAspHisArgValProArgIleTyrValProLeuLysA 282  
 :  
 1200 CTTTGTGCGCGCGCGTAAGATTCCGCGATCCGATCGAAGAACGCGAC 1249  
 282 sPCysProGlnAspPheMetThrArgProLysAspPheAlaAsnThrLeu 298  
 :  
 1250 AGCGAGCAATG.....CTGCAGAACCGCGC 1275  
 299 PheIleCysArgIleIleAspTrpLysGluAspCysAsnPheAlaLeuG 315  
 :  
 1276 ATCATCGTAACGATTGACATGGCGGCACTCGCGCTACCCACATGG 1325  
 315 yGlnLeuAlaLysSerLeuGlyGlnAlaGlyIleGluProGluThrG 332  
 :  
 1326 ACATTTTGTGGCTCCCTGGACCGCTCGAGACATGCCACCGAGACG 1375  
 332 luGlyIleLeuThrGluTyrGlyValAspPheSerAspPheSerSerGlu 348  
 || :  
 1376 AGGTTATCTCTGTGGAGCAGCTCCCGCACTGCAAGTTCTCGGACGAA 1425  
 349 ValLeuGluCysLeuProGlnSerLeuProTyrThrIleProProaspG 365  
 :  
 1426 GTGCTTACCTTCTGCCGAA...ATGCCCTGGACAATTACTGACGAGGA 1472  
 365 uValGlyLysArgAspLeuArgLysAspCysIlePheThrIleasp 382  
 :  
 1473 CTATTCGAAGCGTGTGGATCTGAGGATCTGTACATTTGCTCTGTGGATC 1522  
 382 rSerThrAlaArgAspLeuAspAlaLeuAlaCysArgArgLeuThr 398  
 :  
 1523 CCGCGGCTGCACGGATATCGATGATGCTTTCACCTGCAAGGAACCTACC 1572  
 399 AspGlyThrPheGluValGlyValHisIleAlaAspValSerTyrPheVa 415  
 :  
 1573 AACGGCAATCTGGAAGTGGGCTTCATAGCCGACGTAAGTCACTTTAT 1622  
 415 lProGluGlySerSerLeuAspLysValAlaGluArgAlaThrSerV 432  
 :  
 1623 TCGTCCAGGAACGATGACATGGAGCGCGCGGCGGAGGACACCG 1672  
 432 alTyrLeuValGlnLysValValProMetLeuProArgLeuLeuCysGlu 448  
 ||||| :  
 :





550 spGlnLeuLysLeuAlaPheThrLeuAspHisGluThrGlyLeuProGln 566  
::: ||||| ||| ||||| ||||| ||||| |||||  
3352 ACAATGCCAAGCTACGCTTCTCTCGATCCAAATTACTGGCGAGCGGTG 3401  
::: ||||| ||||| ||||| ||||| ||||| |||||  
567 GlyCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPhe 583  
::: ||||| ||||| ||||| ||||| ||||| |||||  
3402 TCTTTTGAAGTGGAGAAACAGCAGAGGCTAACCGGCTGATTGAAGATT 3451  
::: ||||| ||||| ||||| ||||| ||||| |||||  
583 eMetLeuLeuAlaAsnMetAlaValAlaHisIlePheArgThrPheP 600  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
3452 FATGCTCTGCCAACCGAGCAGTCGCCGCTTTATACAGACTCCTTTTC 3501  
::: ||||| ||||| ||||| ||||| ||||| |||||  
600 roGluGlnAlaLeuLeuArgHisProProGlnThrLysMetLeu 616  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
3502 CGGATATCGTGTACTGCTGTAACACCCCTCCGCCACTTATTAAGTCCCTT 3551  
::: ||||| ||||| ||||| ||||| ||||| |||||  
617 SerAspLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSe 633  
::: ||||| ||||| ||||| ||||| ||||| |||||  
3552 AAGGCTTTAAGGNAAGCTGCTGCTGGGATTCGAATTGGACTATAG 3601  
::: ||||| ||||| ||||| ||||| ||||| |||||  
633 rSerAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspL 650  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
3602 CTCGCTCAAGCGCTCCAGAGAGCATGTCGGTTGTGCAATGAGCGC 3651  
::: ||||| ||||| ||||| ||||| ||||| |||||  
650 ysTyrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgPro 666  
::: ||||| ||||| ||||| ||||| ||||| |||||  
3652 CCAATCTCTGCGCAATGAATGCTGCTAAGTCAGCTTTTGATGAACCA 3701  
::: ||||| ||||| ||||| ||||| ||||| |||||  
667 MetGlnMetAla 670  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
3702 ATGGCGCGTGCACAGTGAAGTCTCAGTTTCAAATAATTACTCTTCCCA 3751  
::: ||||| ||||| ||||| ||||| ||||| |||||  
671 3752 CAATTTCCCTTTAGATATTTGTAGCGAGGTAATCGAACCAGCGC 3801  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
682 uGlnPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrS 699  
::: ||||| ||||| ||||| ||||| ||||| |||||  
3802 TGACCTGTGGCAGTACGCTTGTCCATACCCATATACACTCACTTTACTA 3851  
::: ||||| ||||| ||||| ||||| ||||| |||||  
699 erProIleArgArgPheAlaAspValIleValHisArg 711  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
3852 GCCCAATCCGTCGTATCTCGATATTTGGTGACCGGTAACGTGCACAT 3901  
::: ||||| ||||| ||||| ||||| ||||| |||||  
712 3902 ACCTTGGAAAGTTCGCTCCCGTACTTAAATTTCCCTCCACTCTAGACTTT 3951  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
713 euAlaAlaLeuGlyTyrSerGluGlnProAspValGluProAspThr 729  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
3952 TGGCGCGGCACCTAAATATTGTACACGCGCTAAACGTACCAGATGAT 4001  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
730 LeuGlnLysGlnAlaAspHisCysAsnSprArgMetLaserLysAr 746  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
4002 TTGCACACTTACGAAATAGCAATGAGCGCAAGTACAAATCGCAAAA 4051  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
746 gValGlnGluLeuSerIleGlyLeuPheAlaValLeuValLysGluS 763  
::: ||||| ||||| ||||| ||||| ||||| |||||  
4052 GGCGGAGAGGAGGACTCTGTGTAACCTTCAACAGATATGTTCAACA 4101  
::: ||||| ||||| ||||| ||||| ||||| |||||  
763 erGlyProLeuGlnSerGluAlaMetValMetGlyValLeuAsnGlnAla 779  
::: ||||| ||||| ||||| ||||| ||||| |||||  
4102 AGCAGGCAATATACATGCGAGCTGTGGTCATTTGAATATTTTTCAGCAT 4151  
::: ||||| ||||| ||||| ||||| ||||| |||||  
780 PheAspValLeuValLeuArgPheGlyValGlnLysArgIleTyrCysAs 796  
::: ||||| ||||| ||||| ||||| ||||| |||||  
4152 ATGAACGTAGTTACTCTGGAGTCGGGC 4180  
::: ||||| ||||| ||||| ||||| ||||| |||||  
796 nAlaLeuAlaLeuArgSerTyrSerPheGlnLysVal 809  
::: ||||| ||||| ||||| ||||| ||||| |||||  
4181 CGTTATGAAGCATTTAACTACAAATGCAAAAGGTCCTCGTAGACAC 4227

809 lyLysLysProGluLeuThrLeuValTrpGluProAspAspLeuGlu 825  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
4228 ACGCGTACCACCAATTACATTCTTATTCGCGAG...CGCAACTTAAGCAG 4274  
::: ||||| ||||| ||||| ||||| ||||| |||||  
826 GluProThrGlnGlnValIleThrIlePheSerLeuValAspValVal 842  
::: ||||| ||||| ||||| ||||| ||||| |||||  
4275 TCACCTAGGAG.....TTCCAGCTCCTTTCGTCGTGCCCATATGTCT 4318  
842 u 842  
|  
4319 G 4319

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL11058

seq\_documentation\_block:

ID ABL11058 standard; cDNA; 5347 BP.

XX ABL11058;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27656.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66955.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Claim 1; SEQ ID NO 27656; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 5347 BP; 1410 A; 1280 C; 1347 G; 1310 T; 0 other;

alignment\_scores:

Quality: 920.50 Length: 916

Ratio: 1.757 Gaps: 28

Percent Similarity: 57.205 Percent Identity: 29.476

alignment\_block:

US-09-647-377-9 x ABL11058 ..

Align seg 1/1 to: AB111058 from: 1 to: 5347

```

20 SerSerValValGlyProSerAlaValGlyAlaSerProGlyAsp.LysL 36
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1712 AGCAGCGGAGTACGTGAATCTCTCAGAGATTTTCCACTTCTGATGGATA 1761

36 ySerLysAsnLysSerMetArgGlyLysLysSerIle.....Phe 50
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1762 ACCTGTCGCAAAACAAATTCGACAGCAGAGAAGATGGTCTGCCCCAGTAC 1811

51 GluThrTyrMetSerLysGlyAspValSerGluGlyLeuLysArgGlyTh 67
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1812 CGCGCGCATCTCAGCATGAAGAGCTGCTGGAAGACTGCGCCAAAATAA 1861

67 IleuIleGlnGlyValLeuArg..... 74
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1862 GTACTACAGGCGACTTTTCAGCATCCAGGAAACACTTGTGAGGGCA 1911

75 ..IleAsnProLysPheHisGluAla.....Phe 84
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1912 CGTCAATCTAGAGAAATTCGAAAAGGGGTAAAGTTAAACGCTAAATAT 1961

85 IleProSerProAspGlyAspArgAsp.....IlePheI 96
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1962 CTTCCTGGCATACGCTTCACATACGATTTCTGTTTAATAGATCCTTAT 2011

96 eAspGlyValValAlaArgAsnArgAlaLeuAsnGlyAspLeuValV 113
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2012 CCAAGGCGCGAGTCTCTTAATCGCGCGTGGATGGCACCCTGGTGCGCG 2061

113 allLysLeuLeuProGluAlaThrTriPysAlaValLysProGluSerAsn 129
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2062 TGGAGCTTTGCCGAGCGGGAGTGTGTACGA.....CCAGTGAAT 2105

130 AspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGlyCy 146
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2106 GTACTGAGGAGAGAATGCTATGCGCGATGAAGTGCC..... 2144

146 sGlyHisProLeuGlnSerArgLysGlyTrpSerGlyProAspV 163
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2144 ..... 2144

163 allIleGluAlaGlnPheAspSerAspSerGluAspArgHisGly 179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2145 .....AGCGAGAAAGGCCCAAG 2162

180 AsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerThrPr 196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2163 GATGAANAAGAGATGTTGAATCAGTGGCTGCGCTGCTTATCGCA.. 2210

196 oAspArgGlyLysGluAspSerSerThrProValMetLysAspGluAsn 213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2211 .....GAGCGCA 2217

213 hrProIleProGlnAspThrArgGlyLeuSerGluLysSerLeuGlnLys 229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2218 CGCCC..... 2222

230 SerAlaLysValValTyrIleLeuGluLysLysHisSerArgAlaAlaTh 246
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2223 ACAGGGCAATTTGTCGGTATTGTCGCGACGGAAATGG...CGACAGTATTG 2269

246 rGlyIleLeuLys.....LeuLeuAlaAspLysAsnSerAspLeuPheL 261
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2270 CGGCATCTGCGACCTAGCTTTAATAGAGGATACAAATCGGCAC..... 2312

261 yLysTyrAlaLeuPheSerProSerAspHisArgValProArgIleTyr 277
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2313 .....ATCTTTGTCGCGCGACCGTAAGATTCGGGATCCGCTCCGT 2351

278 ValProLeuLysAspCysProGlnAspPheMetThrArgProLysAspPh 294
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2352 ATCGAAACGCGACAGCGCAATG.....CT 2377

```

```

294 eAlaAsnThrLeuPheIleCysArgIleIleAspTrpLysGluAspCysA 311
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2378 GCAGAACCCAGCATCATCTGAAGATTGACACATGCCGCGCAACTGCC 2427

311 snPheAlaLeuGlyGlnLeuAlaLysSerLeuGlyGlnAlaGlyGluIle 327
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2428 GCTACCCACATGGACATTTTGTGGCTCCCTGGGACCGCTCGGAGACATG 2477

328 GluProGlnThrGluGlyIleLeuThrCluTyrGlyValAspPheSerAs 344
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2478 GCCACCAGAACGAGTTATCTCTGGAGCAGCAGCTCCCGACTGCAA 2527

344 pPheSerSerGluValLeuGluCysLeuProGlnSerLeuProTrpThrI 361
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2528 GTTCTCGGACGAAGTCTTAGCTTCTGCCGAAA...ATGCCCTGGACAA 2574

361 leProProAspGluValGlyLysArgArgAspLeuArgLysAspCysIle 377
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2575 TTACTGACGAGGACTATTCGAAGCGTGGATCTGAGGATCTGTACATT 2624

378 PheThrIleAspProSerThrAlaArgAspLeuAspAlaLeuAlaCy 394
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2625 TGCTCTGTGATCCCGGGGTGCACGATATCATGATGCTTTGCACITG 2674

394 sArgArgLeuThrAspGlyThrPheGluValGlyValHisIleAlaAspV 411
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2675 CAAGAACTACCGAAGCGCAATCTGGAAGTGGGCTTTCATCATAGCCGAG 2724

411 alSerTyrPheValProGluGlySerSerLeuAspLysValAlaLaGlu 427
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2725 TAAGTCACTTATTCGTCAGGAAACGCACATGGACATGGAGCGCGCC 2774

428 ArgAlaThrSerValTyrLeuValGlnLysValValProMetLeuProAr 444
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2775 AGGGACACCGCTGACTGTGGCAACCAATCCACATGATGTTACCGCA 2824

444 gLeuLeuCysGluGluLeuCysSerLeuAsnProMetThrAspLysLeu 461
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2825 GTTACTCAGCTCAAAATCTCTGCTCATTTGTTGGCGGTGTGAGCGATTG 2874

461 hrPheSerValIleTrpLysLeuThrProGluGlyLysIleLeuGlu 477
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2875 CTTCTCTCGTGTGGAGTGGCAATGAAGCAAAATGTTTAAGCAAA 2924

478 TrpPheGlyArgThrIleIleArgSerCysThrLysLeuSerTyrAspHi 494
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2925 CGCTTCCACAAGAGCGCTCATCAAAATCGAAGCGTGCCATGACCTACGAGA 2974

494 sAlaGlnSerMetIleGluAsnProThrGluLysIleProGluGluL 511
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2975 AGCGAGAAATATCATGACGATGCCACGCAACAG..... 3008

511 euProIleSerProGluHisSerValGluGluValHisGlnAlaVal 527
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3009 .....AATGAATAGCAAGTCCCTG 3029

528 LeuAsnLeuHisSerIleAlaLysGlnLeuArgArgGlnArg..... 541
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3030 AGAATCTTAATCGTTGGCAAAATATCTGAAAAGCGACGATGATGA 3079

541 ..... 541

3080 TGGGTAAAGTTATATTAAACCTCGTAAATCAAGTGAACTAAGGCATCCAT 3129

542 ..PheValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaPheThr 557
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3130 AATTTACAGA...GCCCTTGCTTGCCTCCCGGAGATCCGCTTCCAG 3175

558 LeuAspHisGluThrGlyLeuProGlnGlyCysHisIleTyrGluTyrAr 574
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3176 GTGGACGAGACACACGAACCACTCGAGGTGGAGGTGAAGCAGATGCC 3225

```



```

99 a1.....ValAlaArgAsnArgAlaLeuAsnGlyAspLeuVal 111
   :: |||||::: ||::: ||||| ||
1034 TCGTCTCTTCCAGTGGCTAGCGCTAGCTCGCTCGATGGTGCACAAAGTA 1083
112 ValValLysLeuLeuProGluAspGlnTrpLysAlaValLysProGluSe 128
1083 ..... 1083
128 rAsnAspLysGluIleGluAlaThrTyrGluAlaAspIleProGluGlu 145
   ::|||::: ||::: ||
1084 .....CGTGCATTCTGCTAAACCCCTG 1105
145 LysCysGlyHisHisProLeuGlnGlnSerArgLysGlyTrpSerGlyPro 161
   ||
1106 GT..... 1107
162 AspValIleIleGluAlaGlnPheAspAspSerAspSerGluAspArgH1 178
   |||||::: |||||
1108 .....GCTCAGGTAGTTCAAAGACAGCTGAGCCCTCGTC 1142
178 sGlyAsnThrSerGlyLeuValAspGlyValLysLysLeuSerIleSerT 195
   ||::: ||||| ||::: |||||:::
1143 TGTGAATCTCTGGT.....GGAGCCCTCATTTATCTTTATAT 1183
195 hrProAspArgGlyLysGluAspSerSerThrProValMetLysAspGlu 211
   :: |||
1184 CGTCGGAC..... 1191
212 AsnThrProIleProGlnAspThrArgGlyLeuSerGluLysSerLeuG1 228
   || |||
1192 ACTGCCCT..... 1200
228 nLysSerAlaLysValValTyrIleLeuGluLysLysHisSerArgAlaA 245
   ::|||::: |||::: ||::: ||
1201 ....AAAGCGTTCGTATGCAATAATAACCAACTAGCTCGGCCAGA 1246
245 leThrGlyIleLeu.....LysLeuLeuAlaAspLys 255
   |||||::: |||||
1247 TCGTGGGAACCATATCGTTTACGAACCCAACTCTGTGACGACCAG 1296
256 AsnSerAspLeuPheLysLysTyrAlaLeuPheSerProSerAspHisAr 272
   |||||::: |||||::: |||||
1297 .....CTATTCTACAAA.....TTCCGACCGCTACGATTTGCG 1328
272 gValProArgIleTyrValProLeuLysAspCysProGlnAspPheMetT 289
   |||||::: |||||::: |||||
1329 TGTCCCATGGTCTACGTTCCGAAGGATGCTGTGCCGCCACATCGGAA 1378
289 hrArgProLys...AspPheAlaAsnThrLeuPheIleCysArgIleIle 304
   ::|||::: |||::: |||||::: |||||::: |||||
1379 ACAAGCAGCAATAGACGTGTCGGGCTCTTATACCTTGGCCCATATACTC 1428
305 AspTrpLysGluAspCysAsn...PheAlaLeuGlyGlnLeuAlaLysSe 320
   |||||::: |||||::: |||||
1429 GAG.....ACAGATTGCAATGGGCACATCGCTGAACTAATCCAGCC 1472
320 rLeuGlyGlnAlaGlyGluIleGluProGluThrGluGlyIleLeuThrG 337
   ::|||::: |||||::: |||||
1473 GGTGGCCGCGTGGTAAATTGGATGACGAACCTTAAGGCAATTCGTTC 1522
337 lutyrglyVal...AspPheSerAspPheSerSerGluValLeuGluCys 352
   |||||::: |||||::: |||||
1523 ATAACGGTCTCGGGATATAAGCCATTGAGCAACGATTTATCGACATT 1572
353 LeuProGlnSerLeuProTrpThrIleProProAspGluValGlyLysAr 369
   ||| ||| |||
1573 TACTCCAGCCGCCACCT...CCGATAAGCCAAAGAGGATCTCGGTCAACG 1619
369 gArgAspLeuArgLysAspCysIlePheThrIleAspProSerThrAlaA 386
   ::|||::: |||||::: |||||
1620 AAAAGACTTAAGAAAGATGTCATATTTACTATAGATCGACTGCTC 1669
```

---

```

386 rgAspLeuAspAspAlaLeuAlaCysArgArgLeuThrAspGlyThrPhe 402
   |||||::: |||||::: |||||::: |||||::: |||||
1670 GCGATTGGATGACGCGGTTTCTATAGAGAGCTAGGCGACACGAATAT 1719
403 GluValGlyValHisIleAlaAspValSerTyrPheValProGluGlySe 419
   ||::: |||||::: |||||::: |||||::: |||||
1720 GAGATTGGCTCCATATATCCGATGTTTCGCACTTCTGTATAGAGGACAA 1769
419 rSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyrLeuValG 436
   ::|||::: |||||::: |||||::: |||||::: |||||
1770 TGAACCTGGATACATAGTAGGAAGACGCTTACGCTCTATTATTAGCCA 1819
436 InLysValValProMetLeuProArgLeuLeuCysGluGluLeuCysSer 452
   |||||::: |||||::: |||||::: |||||
1820 ACGAGGTAATCCACATGCTGCCTCAGTCTCTATGCATGCGA...TGTTC 1866
453 LeuAsnProMetThrAspLysLeuThrPheSerValIleTrpLysLeuTh 469
   ||| ||| |||
1867 CTGTCCTCGGCAAGACAAAGTTTCTTTTCCGCTCTTCTGGCGGATGAA 1916
469 rProGluGlyLysIleLeuGluGlu.....TrpPheGlyArgThrIleI 484
   ::|||::: |||||::: |||||
1917 CGGGAAGGGAGTCACTGCTGCAAAAGAACCCAGAATTTTGCACACAGTCA 1966
484 leArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGlu 500
   ||::: |||||::: |||||::: |||||::: |||||
1967 TTAACCTCTGCTCGCAATTTGCTTATGAACACGCTCAAAAGATCATCGAC 2016
501 AsnProThrGluLysIleProGluGluLeuProProIleSerProG1 517
   |||||::: |||||::: |||||
2017 AACCTAATGAACGGTTTACCAGAACGACTTTCGCCACCATCTCGAATGG 2066
517 uHisSerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleA 534
   ::|||::: |||||::: |||||
2067 ATTTAATCCCGATGATCAGAAATAGGTGCTATGCTGCACGACATG 2116
534 laLysGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAsp 550
   |||||::: |||||::: |||||
2117 CAAGCTCATCTCCGTAAGACACGCTTTAGACAATGGCGCACTCAACAATA 2166
551 GlnLeuLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnG1 567
   ::||| ||||| |||
2167 ATGCCAAGTACGCTTCTGCTCATCCATTACTGGCGAGCGGTGTC 2216
567 yCysHisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPheH 584
   ::|||::: |||::: |||||::: |||||
2217 TTTTGAAGTGGAGAAACAGCAGAGAGGCTAACCGGCTGATTGAAGAGTTTA 2266
584 etLeuLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPhePro 600
   |||||::: |||||::: |||||
2267 TGTCTCGGCCAACAGGAGGCTGCGCCCTTTATACAGCACTCCTTCGCG 2316
601 GluGlnAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSe 617
   ::|||::: |||||::: |||||
2317 GATATCGCTGCTACTGCTTAACCAACCGCTCCGCCACTTATTAAAGTCCCTAA 2366
617 rAspLeuValGluPheCysAspGlnMetGlyLeuProMetAspValSerS 634
   ::||| ||||| ::|||::: |||||
2367 GCTTTAAGGAAAGCTGCTTCTCTGGGATTCGAATTCGACTATAGCT 2416
634 erAlaGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspLys 650
   |||||::: |||||::: |||||
2417 CGTCTAAGCGCTCCAGGAGAGCATGCTGCGGTGTGTGCAATGAGGCGCC 2466
651 TyrSerLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMe 667
   ::|||::: |||||::: |||||
2467 AATCTGTGCGCATGAATGCTGCTGCTAGTCAGCTTTTGATGAACCAAT 2516
667 tGlnMetalLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluGlnP 684
   ||| |||
2517 GCGCCGCTGCA..... 2526
684 heArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerPro 700
```

```

2526 ..... 2526
701 IleArgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLe 717
2527 .....ACACTTTTGGCGCCGCACT 2546
717 uGlyTyrSerGluGlnProAspValGluProAspThrLeuGlnLysGlnA 734
2547 TAAATATTGTAACCGCTAAAGTACCACAGATGATTGCACACCTTAA 2596
734 IaAspHisCysAsnAspArgMetAlaSerLysArgValGlnGluLeu 750
2597 CAAATTAGCGATGAGCCAGTCAATGCGAAAGCCGAGAGGAC 2646
751 SerIleGlyLeuPhePheAlaValLeuValLysGluSerGlyProLeuG 767
2647 TCTGTAACTGTACTTCAACAGATATGTCACAACAAGCAGGCATATA 2696
767 uSerGluAlaMetValMetGlyValLeuAsnGlnAlaPheAspValLeu 784
2697 CATCGAGCTGTGTCATTGAATATTCAGCATATGATGAACCTAGTTA 2746
784 aLeuArgPheGlyValGlnLysArgIleTyrCysAsnAlaLeuAlaLeu 800
2747 CTCGGAGTCGGC.....CACGTTATAAGCATT 2775
801 ArgSerTyrSerPheGlnLysVal.....GlyLysLysProG 813
2776 ...AACTACAAATGCAAAAGTCTCTGTAGACACACGCGCTACCCAA 2822
813 uLeuThrLeuValTrpClnProAspAspLeuGluGluProThrGlnG 830
2823 TTACATTCTTATGCCGAG...CGCAACTTAAAGCAGTCACCTAGGAAG. 2868
830 lnValIleThrIlePheSerLeuValAspValValLeu 842
2869 .....TTGCAGCGCTTTCGGTGGTCCCATATGCTGTG 2901

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.AAS79116
seq_documentation_block:
ID_AAS79116 standard; cDNA; 4309 BP.
AC AAS79116;
XX
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #14920.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB; ABG14929.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT

```

PT responsible for genetic disorders or other traits and to assess biodiversity -

XX Claim 1; SEQ ID No 14920; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4309 BP; 1450 A; 937 C; 961 G; 961 T; 0 other;

alignment\_scores: Quality: 616.50 Length: 814  
Ratio: 1.577 Gaps: 26  
Percent Similarity: 48.034 Percent Identity: 24.693

alignment\_block:

US-09-647-377-9 x AAS79116 ..

Align seg 1/1 to: AAS79116 from: 1 to: 4309

```

93 AspIlePheIleAspGlyValValAlaArgAsnArgAlaLeuAsnGlyAs 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
643 GACATCTTAATCCACGGGATGAAGCTCGAAACCGCTCAATTCATGAGA 692
109 pLeuValValLysLeuLeuProGluAspGlnTrpLysAlaValLysP 126
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
693 TGTGTTAGTTGTGGAGCTCTCTCTAAATGAATGAAGGA..... 735
126 roGluSerAsnAspLysGluIleGluAlaThrTyrGluAlaAspIlePro 142
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
736 .....AGAACCGTAGCCCTGTGTGAGAATGACTGTGAC 768
143 GluGluGlyCysGlyHisHisProLeuGlnGlnSerArgLysGlyTrpse 159
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
769 GACAAGGCTTCGGGAGTCCCA..... 792
159 rGlyProAspValIleIleGluAlaGlnPheAspSerSerAspSerClua 176
792 ..... 792
176 spArgHisGlyAsnThrSerGlyLeuValAspGlyValLysLeuSer 192
792 ..... 792
193 IleSerThrProAspArgGlyLysGluAspSerSerThrProValMetLys 209
792 ..... 792
209 sAspGluAsnThrProIleProGlnAspThrArgGlyLeuSerGluLys 226
:::|:|:|:|:|:|
793 .....AGTGAGCCCATGCT..... 807
226 erLeuGlnLysSerAlaLysValValTyrIleLeuGluLys..... 239

```

```

808 .....:||||| |||||:||||
240 .....ACAGTCCAGTGGTGGCATACTTCAGAAAGAACTGGCGG 846
240 .....LysHisSerArgAlaAlaThrGly1 248
847 GATTATGTGGTGACATTTCCTGCCAAAGAGAGGTCCAATCTCAGGCGC... 894
248 eLeuLysLeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaL 265
895 .....AAAAATGCT.....CAGAAATCC 913
265 eUpHeSerProSerAspHisArgValProArgIleTyrValProLeuLys 281
914 TGGTTACACCTTGGGATTACAGAAATTCACAAAATTCGAATTAGCACTCAG 963
282 AspCysPro.....GlnAspPheMetThrArgProLysAspPheAlaAs 296
964 CAAGCAGAAACCCCTCCAGGACTTCAGGGTG..... 993
296 nThrLeuPheIleCysArgIleLeaSpTrpLysGluAspCysAsnPhea 313
994 .....GTCGTGGCATCATCTCTGGGAGTCAACATCTGTGTATC 1033
313 lAlaLeuGlyGlnLeuAlaLysSerLeuGlyGlnAlaGlyLeuIleClnPro 329
1034 CAATGGCACTTTTGTGGTGTGTAGGAAGATCGGAGATCTCGAAGGG 1083
330 GluThrGluGlyIleLeuThrGluTyrGlyValAspPheSerAspPheSe 346
1084 GAAATTGCAACCATCTCGTGGGAAACAGTATTTCAGTTATTCCTTTC... 1131
346 rSerGluValLeuGluCys.....LeuProGlnSerLeup 358
1132 .TCAGAGCTCAGATGTGTGAGATGCCATAACACACCCAGAAAGT...C 1177
358 roTrpThrIleProAspGluValGlyLysArgAspLeuArgLys 374
1178 CDTGGAAGTGAGTCTCTGAGAGGAGCAAAAAACGTTAAAGACTTGAGGAA 1227
375 Asp...CysIlePheThrIleAspProSerThrAlaArgAspLeuAspAs 390
1228 AGCATCTCGATATTCAGATGTGACCCCAAGGTTGTGAAGATGTGGATGA 1277
390 pAlaLeuAlaCysArgArgIleuThrAspGlyThrPheGluValGlyValH 407
1278 CACACTCTCAGTCAGAACCTTAATAATGCAACCTGGAACTTGGGGTCC 1327
407 iStleAlaAspValSerTyrPheValProGluGlySerSerLeuAspLys 423
1328 ACATCCAGATGTACACACTTGTGGCACCACCAANTTCTACATTTGATATT 1377
424 ValAlaAlaGluArgAlaThrSerValTyrLeuValGlnLysValValPr 440
1378 GAAGCTAGAACCAAG.....TATGCTGTA..... 1401
440 oMetLeuProArgLeuLeuCysGluGluLeuCysSerLeuAsnProMetT 457
1401 ..... 1401
457 hrAspLysLeuThrPheSerValIleTrpLysLeuThrProGluGly... 472
1402 .....AGCATCATGTGGGAACGTGATAAAGCCTCTTAT 1434
473 LysIleLeuGluGluTrpPheGlyArgThrIleIleArgSerCysThrLy 489
1435 GAAATTAGAAAGTGTGTATGGCAGAACCATTTATTCGATCAGATACAA 1484
489 sLeuSerTyrAspHisAlaGlnSerMetIleGlu.....AsnProt 503
1485 ACTGTTCTATGAGCAGCCCAAGAACTACTGGATGGAAACTTAAGCGTTG 1534
503 hrGluLysIleProGlu...GluGluLeuProIleSerProGluHis 518
:||||| :|||

```

```

1535 TTGATGATATTCCAGAATTCAAAGACTTGGATGAGAAAGAGCAGACAAGCC 1584
519 SerValGluGluValHisGlnAlaValLeuAsnLeuHisSerIleAlaLy 535
1585 AAGCTGGAGAGTGTGTGGGCAATTTGGAAGCTGACCGACATAGCTCG 1634
535 sGlnLeuArgArgGlnArgPheValAspGlyAlaLeuArgLeuAspGlnL 552
1635 CCATGTCAGAGCTAAACAGCAGCGGATGTGTGCTGCTGGAACCTGGAAGGG 1684
552 eUlsLeuAlaPheThrLeuAspHisGluThrGlyLeuProGlnGlyCys 568
1685 TAGAGTTTTCGTACAGCTAGATGACAAAAGAACATT..... 1722
569 HisIleTyrGluTyrArgAspSerAsnLysLeuValGluGluPheMetLe 585
1722 ..... 1722
585 uLeuAlaAsnMetAlaValAlaHisLysIlePheArgThrPheProGluG 602
1723 .....C 1723
602 lAlaLeuLeuArgArgHisProProGlnThrLysMetLeuSerAsp 618
1724 ACGACCTCATCCCAAGCAGCGCCGGAAGTCCAATAAAACACTGGCTGAT 1773
619 LeuValGluPheCysAspGlnMetGlyLeuProMetAspValSerSerAl 635
1774 TCTCTGGAT..... 1782
635 aGlyAlaLeuAsnLysSerLeuThrLysThrPheGlyAspAspLysTyrS 652
1783 .....AATGCGAAGCAGCCCCACG 1801
652 erLeuAlaArgLysGluValLeuThrAsnMetTyrSerArgProMetGln 668
1802 ATCCCATTTGAAACAGGCTACTGCTCCATGCGCCAGCAGGACCGATCG 1851
669 MetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluGlnPhear 685
1852 AATGCTGTGTACTTCTCCACCGGATCCTGTGCG...GAGGAGGATTC 1898
685 gHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIleA 702
1899 TCATTACGGCTTGCATTAGATAAATATACCCTTACTTCTCCATRA 1948
702 rArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGly 718
1949 GAAGATATTGATATTGTAGTACACCGCTTGTATGTCAGCATTC 1998
719 TyrSerGluGlnProAspValGluProAsp.....ThrLe 730
1999 AAGATAGAAATGGAATTAAGGAATCTGTTCCAGCAACAAGATCT 2048
730 uGlnLysGlnAlaAspHisCysAsnAspArgMetAlaSerLysArgV 747
2049 TGAGGAATTTATGCAGACATATCAACAACAGAAACCAAGCAGCAGCAT 2098
747 aGlnGluLeuSerIleGlyLeuPhePheAlaValLeuValLysGluSer 763
2099 CTCAGAACGACTCTACTGAGCTCTTCCAGTGTACTTCAAGACAAA 2148
764 GlyProLeuGluSerGluAlaMetValMet...GlyValLeuAsnGlnAl 779
2149 GACCTGCGCCAGGAGGCTTGCATATCTGACGAGGTATT..... 2190
779 aPheAspValLeuValLeuValLeuValGlnLysArgIleTyrCysA 796
2190 ..... 2190
796 snAlaLeuAlaLeuArgSerTyrSerPheGlnLysValGlyLysLysPro 812
2191 .....TATTCAATTAGAACAAATGGTAATAA... 2217

```





DE Murine LOBO genomic DNA fragment 2.  
 XX  
 KW LOBO; long bones; bone development; bone extension; skull; osteopathic;  
 KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
 KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.  
 XX  
 OS Mus musculus.  
 XX  
 XX W09950284-A2.  
 PN  
 XX  
 PD 07-OCT-1999.  
 XX  
 XX 26-MAR-1999; 99WO-EF02055.  
 PF  
 XX  
 XX 27-MAR-1998; 98DE-1013799.  
 PR  
 XX  
 XX (ROSE/) ROSENTHAL A.  
 PA  
 XX  
 XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;  
 PI  
 XX  
 XX WPI; 1999-601320/51.  
 DR  
 XX  
 XX Nucleic acids encoding proteins which influence bone development,  
 PT useful for treating and studying bone disorders -  
 PT  
 XX  
 XX Example 3; Page 97-117; 391pp; German.  
 PS  
 XX  
 CC This invention describes novel nucleic acids (I; designated LOBO (long  
 CC bones)) encoding proteins influencing bone development in mammals. The  
 CC proteins of the invention reduce and/or inactivate bone extension (i.e.  
 CC development), with exception of the skull and have osteopathic activity.  
 CC The nucleic acid molecules, proteins and antibodies can be used in  
 CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
 CC and nucleic acid molecules, etc. are useful for production of transgenic  
 CC animals, especially a transgenic mouse for the study of diseases  
 CC associated with bone development, e.g. spondyloepiphyseal dysplasia and  
 CC achondroplasia. This sequence encodes the murine LOBO protein described  
 CC in the method of the invention.  
 XX  
 SQ Sequence 36901 BP; 9597 A; 9409 C; 8973 G; 8922 T; 0 other;  
 alignment\_scores:  
 Quality: 571.00 Length: 906  
 Ratio: 2.813 Gaps: 8  
 Percent Similarity: 22.406 Percent Identity: 22.075  
 alignment\_block:  
 US-09-647-377-9 x AA223892 ..  
 Align seq 1/1 to: AA223892 from: 1 to: 36901  
 667 MetGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGlu1 683  
 15491 CTTCAATGGCACTGACTTCTGCTGGGATGCTGCAGGACCGAGGACGA 15540  
 683 nPheArgHisIstYrAlaLeuAsnValProLeuTyrThrHisPheThrSerP 700  
 15541 GTTCCGGCATTTATGCTCTCAACGTTCCCTCTACACACACATTCACCTCTC 15590  
 700 roileArgArgPheAlaAspValilleValHisArgLeuAlaAlaA 716  
 15591 CCATCCGCCGCTTTGCTGACGCTCATAGTGCACCGCTCTCTGGCTGCTGCT 15640  
 716 ..... 716  
 15641 CTGGGTAAAGGACATGACTCTGGCTGGGAGACCTTTGCTGTCGAGAG 15690  
 716 ..... 716  
 15691 TTACCACCTCTCAGAGTAAGTGACACCATTAAGTATCATGACATGAC 15740  
 716 ..... 716

```
16641 AGAAGGAGGCCAGAGGGTAGCCAGCACTGCTGGACAGTGTGTCCTC 16690
796 ..... 796
16691 AAACAGAAACAAGAGGCCCATCTGCTTGGCTTCTCCCTGGATCAGAG 16740
796 ..... 796
16741 TTATTAACAAGTCTCTTGGTGGCTTCTGTATGTCACTTGGGGGGCTTT 16790
796 ..... 796
16791 GCTTTAGTGTCTCTGTGTCAACAAGTCAACACCTGGCTCTACCCCTGG 16840
796 ..... 796
16841 CTTTCAACTTCTTACATACACTTGGGGAAGTGTGAACCTTGCACGAA 16890
796 ..... 796
16891 GAGACACAGGATTCATGAAGAGGCAGAACAGGAAAGGCCAAGTGCAAG 16940
796 ..... 796
16941 TGGAACTACCAACACCTGTAGTTACCTGGCTCTCAGCCTGGTGGTCAGG 16990
796 ..... 796
16991 TCTATCACCACAGCCTAGGAGATCTCTTCTTTTGTCTACAGTCAACCAC 17040
796 ..... 796
17041 CCTCCACATGTCCCTTGGAAATGGGTCACTTCAGGTTCTACTTTGAC 17090
796 ..... 796
17091 CAAAGTGCTATTAGCAGAACCTCTCTAAATCTGGCTGAGTGGACCAAGGA 17140
796 ..... 796
17141 TAGGGGGCTGGGGATGCTCTGTCCAAAGCAGCAGTACAGTAAGGCAG 17190
796 ..... 796
17191 CCGGTACAAAGCTCCCTCCAGCCAGTCAAGAAATAGCAGCAGGAGGAGAA 17240
796 ..... 796
17241 GAGGTCTGTGAAGCCCATAGCCTGAGGCTCCGGTGTGTCCTCCCTGCCCTC 17290
797 Ala..LeuAlaLeuArgSerTyrSerPheGlnLysValGlyLysLysPro 812
||||| :|||
17291 AGGCATGGCCCTGGGATTCCTACAGCTTCAGAAAGTGGGGAAGAGCCA 17340
813 GluLeuThrLeuValTrpClnProAspLysLeuGluGluGluProThrG1 829
||||| :|||
17341 GAGCTCACTCTGTGTTGGAGCCTGTGACCTTGAAGAGAGGCCAACACA 17390
829 nGlnVal.Ile..... 832
||||| :|||
17391 GCAGGTGATCCCTCCCTGCTGTCTCCCTAAGCCTACCTCTGCTCAAAAGTG 17440
832 ..... 832
17441 TGCCCCCTAGTCTCATCTGCCCTCATTTCTCCCCAGCAGCATAGTTCC 17490
832 ..... 832
17491 CCTGTGGGATTCACCAAGCCCTGGCTTAGACTGCCAGGTTCTATATGGG 17540
832 ..... 832
17541 AACACCCCACTATGGCAGTGGTTCTCAACCTTCTGATGACGAGGCCCTTA 17590
```

```
832 ..... 832
17591 ACACAGTTCCATGCTGTGTGACACCCCTTCCCCAGCCATTAAATTAAT 17640
832 ..... 832
17641 TTTGGTTGCTACTTCATATACTATAAGTTTGCTGCTGTATATAATCAAATG 17690
832 ..... 832
17691 TAAATATTTTGGAGATAGAGCAAAAGGCTCTGAACGACAGGTGGGGA 17740
832 ..... 832
17741 CTGCTGCTCTATAGTAGATAGGTGCTATTCTCTCTCCCTGAACAGAACT 17790
832 ..... 832
17791 TTTCAGAAATTTTGAGAGCTCATAAAAGTCTTTTATCCCTCTTGTTC 17840
832 ..... 832
17841 CAAAGGCTGGCCAGCCAGCTCGCGCGGCCAGCCTGTTTCTTGTCTC 17890
832 ..... 832
17891 CTCGTGAATGGTCACGTGAATAACAAATGTCTACATAGTGCCATTAGCCT 17940
832 ..... 832
17941 ACTGGTTTCCCCAGACCAATCCAATCCATTACAGATAGGCGATAGAG 17990
832 ..... 832
17991 GCTCGGGAAGTTAAGTAGCCCTCAGTGTGCTTGGCTTTCATTTGACGC 18040
833 .....Th 833
18041 CCTCACCTGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 18090
833 r1lePheSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuL 850
||||| :|||
18091 CATCTCAGCTGTGTGATGTGTCTCTCAGCAGCAGGCCACACCCCTCA 18140
850 ySTyrSerAlaIleLeuLysArgProGlyLeuGluLysAlaSerAspGlu 866
||||| :|||
18141 AGTACAGTGTCTATCTCAAGCGACCGGCTGGAGAGGCGTCTGATGAG 18190
867 GluProGluasp 870
||||| :|||
18191 GAGCCTGAGGAC 18202

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ23897
seq_documentation_block:
ID AAZ23897 standard; DNA; 38886 BP.
XX
AC AAZ23897;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO homologue genomic DNA fragment 3.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
OS Mus musculus.
XX
PN W09950284-A2.
XX
PD 07-OCT-1999.
```

XX 26-MAR-1999; 99NO-EP02055.  
XX  
XX 27-MAR-1998; 98DE-1013799.  
XX  
XX (ROSE/) ROSENTHAL A.  
XX  
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;  
XX WPI; 1999-601320/51.  
XX  
XX Nucleic acids encoding proteins which influence bone development,  
XX useful for treating and studying bone disorders -  
XX  
XX  
XX Example 3; Page 189-211; 391pp; German.  
XX  
XX This invention describes novel nucleic acids (I; designated LOBO (long  
XX bones)) encoding proteins influencing bone development in mammals. The  
XX proteins of the invention reduce and/or inactivate bone extension (i.e.  
XX development), with exception of the skull and have osteopathic activity.  
XX The nucleic acid molecules, proteins and antibodies can be used in  
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
XX and nucleic acid molecules, etc. are useful for production of transgenic  
XX animals, especially a transgenic mouse for the study of diseases  
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and  
XX achondroplasia. This sequence encodes the murine LOBO protein described  
XX in the method of the invention.  
XX  
XX Sequence 38886 BP; 10087 A; 9920 C; 9356 G; 9523 T; 0 other;

## alignment\_scores:

Quality: 571.00 Length: 906  
Ratio: 2.813 Gaps: 8  
Percent Similarity: 22.406 Percent Identity: 22.075

## alignment\_block:

US-09-647-377-9 x ANZ23897 ..

Align seg 1/1 to: ANZ23897 from: 1 to: 38886

667 MetGlnMetAlaLeuTyrPheCysSerGlyMetLeuGlnAspGlnGluG1 683  
:::|||||||||||||||||||||||||||||||||||||||||||||||||  
17475 CTCAGATGCGACTGTACTTCTGCTCTGGATGCTGCAGGACCGAGGCA 17524  
683 nPheArgHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerP 700  
|||||||||||||||||||||||||||||||||||||||||||||||||  
17525 GTCCGGCATATGCTCTCAAGGTTCCCTCTACACACACTTCACCTCTC 17574  
700 roIlleArgArgPheAlaAspValIlleValHisArgLeuLeuAlaAla 716  
|||||||||||||||||||||||||||||||||||||||||||||||||  
17575 CCATCCGCGCTTTGTGACGTCATAGTGACACCGCTCTGCTGCTGCT 17624  
716 ..... 716  
17625 CTGGGTAAAGGACATGACTCTGGCTGGGAAGACCTTTGCTGTCGAGAG 17674  
716 ..... 716  
17675 TTACCCACTCTCAGAGTAAGTGACCACATTACTGTTATCATCGACATGCC 17724  
716 ..... 716  
17725 GAGGACAGAGAAAGCCTAAGTCTGAACACTGTGATGCCACACCCAGATGA 17774  
716 ..... 716  
17775 TGAAGCTTTTAGTGAGACTTATTGCAAGCGGGGACCATATATGTTCCCA 17824  
716 ..... 716  
17825 GAGCCTTGCCCTCAGCACACAAACCGTCTTATPCCCATACTACTGAAACCCCTG 17874

717 .....Leu..GlyTyrSerGluGlnProAspValGluProAspT 729  
||| |||||||||||||||||||||||||||||||||||||||  
17875 GTGCGCCTCTCTCCAGGCTACAGTAACAGCCAGATGTGGAGCCTGATA 17924  
729 hrLeuGlnLysGlnAlaAspHisCysAsnAspArgArgMetAlaSerLys 745  
|||||||||||||||||||||||||||||||||||||||||  
17925 CCCTACAGAAGCAAGCTGACCACCTGCAATGACCGTCGATGGCTTCCAAA 17974  
746 ArgValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuVal.Lys. 761  
|||||||||||||||||||||||||||||||||||||||||  
17975 CGTGTCCAGGAGCTCAGCATCGCCTCTTCTTCGAGTTCTAGTAAAGGT 18024  
761 ..... 761  
18025 GAGTGTCCAGCCTGCCCCCTTCTTCTTCCCTTCCCTGTCCTCCGATGA 18074  
761 ..... 761  
18075 ATGGAGCACCACTGCAGGTCTCTCCCTGGGAGGATGCCAGATGCTTGT 18124  
762 .....GluSerGlyProLeuGluSerGluAlaMetValMetGlyValL 776  
|||||||||||||||||||||||||||||||||||||||||  
18125 CCTACAGGAGAGCTGSCCCCTTGGAGTCCGAAGCCATGGTGGTGTCC 18174  
776 euAsnGlnAlaPheAspValLeuValLeuArgPheGlyValGlnLysArg 792  
|||||||||||||||||||||||||||||||||||||||||  
18175 TGAACCAAGCTTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18224  
793 IleTyrCysAsn..... 796  
|||||||||  
18225 ATCTACTGCAATGTGAGTATCCCTGGTATGATGATGATGATGATGATGATG 18274  
796 ..... 796  
18275 AGCAAAACCAACCCATTTTCCGCGCTGTGTCTAGTTCTTCCCTTGTGGGA 18324  
796 ..... 796  
18325 AATATTCCCTGGTCCAGAATATCCCATGATAGTATTTACAGGTGTAAATG 18374  
796 ..... 796  
18375 GTGGGATTCACATGAGCTCCCTTCTGTCCCTGGCCATTAGCTATGCAGGG 18424  
796 ..... 796  
18425 CCCACAGACTGCATCTATAGCAGTGAGTTTCACTGGCATGTGCAAGAA 18474  
796 ..... 796  
18475 AGGTCCAGACCCCTGAACCAAGTAGGCTGCCAGGACAGGGCCCTCAG 18524  
796 ..... 796  
18525 GCCAAGGTCAGTCTGAAGTCTTCTTAAAGCCAGGACACTCAGAACA 18574  
796 ..... 796  
18575 TAACAGGATGGCAGGGGTGGGACCTGTGTATGATGTTCTTTATAGAACATGC 18624  
796 ..... 796  
18625 AGAAGGGAGGCCAGAGGGTAGCCAGCACTGCTCTGACACTGTGTCCCC 18674  
796 ..... 796  
18675 AAACAGAAACAGAGGCCATCTGCTTGGCTTCTTCCCTGGATGACAG 18724  
796 ..... 796  
18725 TTTATTCAAAGTCTCTTGTGCTTCTGTAATGTCACTTGGGGGGCTTT 18774  
796 ..... 796

```
18775 GCTTTAGTGTCTGTGGTCACCAAGTCACACCTGGCTCTACCCCTGG 18824
796 ..... 796
18825 CTTTGAACCTCTTACATACACTTGGGGAAGTGTGGAAACCTGCACCTGGAA 18874
796 ..... 796
18875 GAGACACAGGATTCATGAAGAGGCGAGAACAGGAAGGGCCACAGTCAGC 18924
796 ..... 796
18925 TGGAACTACCAGACACCTGTAGTTACCTGGCTCTCAGCCTGGTGTGAG 18974
796 ..... 796
18975 TCTATCACCACAGCCTAGGAGATCTTCTCTTTGCTACAGTCACCAC 19024
796 ..... 796
19025 CCTCCACATTTGTCCTTGGAAATGGGTCACTTCAGGTTCTACTTTGAC 19074
796 ..... 796
19075 CAAAGTGACTTACAGACACCTCTAAATCTGGCTGGAGTGAGGACCAAGGA 19124
796 ..... 796
19125 TAGGGGCTGGGGATGTCTCTGTCCAAGCAGGCAGCTACAGTAAGGCAG 19174
796 ..... 796
19175 CCGGTACAAAGCTCCCTCCAGCCAGTCAGAAATAGCAGGAGGAGAGAA 19224
796 ..... 796
19225 GAGGTGTCTGAAGCCCATAGCCTGAGGCTCGGCTGTCCCTGCCCCC 19274
797 Ala..LeuAlaLeuArgSerTyrSerPheGlnLysValGlyLysPro 812
19275 AGGCACCTGGCCCTGGATCTACAGCTTCAGAGGTGGGAGGAGCAAGCCA 19324
813 GluLeuThrLeuValTrpGluProAspLeuGluGluProThrG1 829
19325 GAGCTCACTCTGTGTTGGGAGCCTGATGACCTTGAAGAGGAGCAACACA 19374
829 nGlnVal.Ile..... 832
19375 GCAGGTCAGTCCCTGTGTGTCCCTAAGCCTACCTCTCTCAACGTG 19424
832 ..... 832
19425 TGCCCTAGGTCCTCATCTGCCTCATTTCTCCAGCACCATAGATTCC 19474
832 ..... 832
19475 CCTGTGGATTCACCAAGCCCTGGCTTAGACTGCCAGGTCTATATGG 19524
832 ..... 832
19525 AACACCCACTATGGCAGTGGTTCACACCTTCTGTATGACGAGCCCTTA 19574
832 ..... 832
19575 ACACAGTCTCTCATCTGTGTGAGACACCTTCCCCAGCACAATAATTAT 19624
832 ..... 832
19625 TTTGCTGTACTTTCATTAACATATAAGTTGCTGCTGTTAATAACAATG 19674
832 ..... 832
```

```
19675 TAAATATTTTGGAGATAGAGCAAGGGTCTCGAACGACAGGTTGGGA 19724
832 ..... 832
19725 CTGCTGCTCTATAGTAGATAGGTGCTATTCTCTCCCTGAAACAGAACT 19774
832 ..... 832
19775 TTTCAGAAATTTGAGAAGCTGATAAAAGCTTCTTTTATCCCTCTGTTC 19824
832 ..... 832
19825 CAAAGGCTGGCCAGCCAGCTCGGCCGCCAGCCTGTTTCTTGTGTC 19874
832 ..... 832
19875 CTCGTGAATGGTCACGTGATTAACAATGCTTACATAGTGCCATTTAGCCT 19924
832 ..... 832
19925 ACTGGTTTCCCGACAGCCCAATGAATCCCATTTACAGATAGGCGATAGAG 19974
832 ..... 832
19975 GCTCGGAAAGTTAAGTAGCCTCAGTGTGCTTGTGATTGCAGGC 20024
833 .....Th 833
20025 CCTCAGCTGCCCTGTCCTCTCTGCTGCTGCTACAGGTATCATC 20074
833 rIlePheSerLeuValAspValValLeuGlnAlaGluAlaThrAlaLeuL 850
20075 CATCTTCAGCCTGCTGTGATGTGCTCTGCAGGCAGAGGCCACAGCCCTCA 20124
850 yTyrSerAlaIleLeuLysArgProGlyLeuGluLysAlaSerAspGlu 866
20125 AGTACAGTGTCTATCTGAAAGCCAGCCCTGGAGAGGCGTCTGATGAG 20174
867 GluProGluAsp 870
20175 GAGCCTGAGGAC 20186
seq_name: /STDs1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT-AAH53049
seq_documentation_block:
ID AAH53049 standard; DNA; 2379 BP.
XX
AC AAH53049;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1491.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US0782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
DR P-PSDB; AAG82199.
XX
```



```
431 SerValTyrLeuValGlnLysValValProMetLeuProArgLeuLeuCy 447
|||||
937 AGTGTGTATCTTGTGATGAGTAAATCCGATGATTCACACCGCTCAAG 986
447 sGluGluLeuCySerLysProMetThrAspLysLeuThrPheSerV 464
|||||
987 TAATGGAATATGCTATTAAATCCAGAGAGATCGTTTAACTTAAGTT 1036
464 alileTrpLysLeuThrProGluGlyLysIleLeuGluGluTrpPheGly 480
|||||
1037 GTCGAATGGAATAATGAACGAGCGCAAGTTGTAAACATGAATCTTT 1086
481 ArgThrIleIleArgSerCysThrLysLeuSerTyrAspHisAlaGlnSe 497
|||||
1087 GATAGTGTATATACATTCAAACTACAGATGACATATGATGCGATTACAA 1136
497 rMetIleGluAsnProThrGluLysIleProGluGlu.....GluLeuP 512
|||||
1137 AATTATCACTGATCAAGATTCTGAATATACGTTTCACAATATAAAGATTAA 1186
512 roPro.....IleSerProGluHisSerValGluGluValHisGln 525
|||||
1187 CACCTATGTAGATTAGCGCAAGATTTATCAATATAGATTATTCAT... 1233
526 AlaValLeuAsnLeuHisSerIleAlaLysGlnLeuArgArgGlnArgPh 542
|||||
1234 .....ATCGCAACACGTCGT... 1248
542 eValAspGlyAlaLeuArgLeuAspGlnLeuLysLeuAlaPheThrLeuA 559
|||||
1249 .....GGAGAAATTTGAT...ATTAATGAAGCGAAAGTACTTG 1288
559 sphisGluThrLeuProGlnGlyCysHisIleTyrGluTyrArgAsp 575
|||||
1289 TGAATGACGAAGGTATTCACAGAGAGTCTAATGAGAGACGTGGCGAA 1338
576 SerAsnLysLeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAl 592
|||||
1339 GGAGAACGTTTAAATGAATCATCTTATAGCAGCAATGAACAGTAGC 1388
592 aHisLysIlePheArgThrPheProGluGlnAlaLeuLeuArgArgHisP 609
|||||
1389 TGAACAC...TTCAATAAATGGAGTAGTACCATTATTATCGTTGTTCAIG 1435
609 roProGlnThrLysMetLeuSerAspLeuValGluPheCysAspGln 625
|||||
1436 AACAAACCAAAATCTGACCGATTAAAGACAGTTCTTCGACTTTATTACCAAT 1485
626 MetGlyLeuProMetAspValSerSerAlaGlyAlaLeuAsnLysSerLe 642
|||||
1486 TTCGGTATT.....AT 1496
642 uThrLysThrPheGlyAspAspLysTyrSerLeuAlaArgLysGluValL 659
|||||
1497 GATAAAAGGTACAGGTGAAGATATTCACCAACACATTACAAAGCATTC 1546
659 euthrAsnMetTyrSerArgProMetGlnMetAlaLeuTyrPheCysSer 675
|||||
1547 AAGAGAAGTTGAAGTAGCAGCAACAAATGGTTATT.....TCAACG 1590
676 GlyMetLeuGlnAspGlnGluGln.....PheAr 685
|||||
1591 ATGATGTTCAGTTCTATCAACAGCACATATGATGATTTAAATTTAGG 1640
685 gHisTyrAlaLeuAsnValProLeuTyrThrHisPheThrSerProIleA 702
|||||
1641 ACATTTGGTTTGTGCTGAGTACTATCTACTCTTACCTTACCTTCAATAC 1690
702 rgArgPheAlaAspValIleValHisArgLeuLeuAlaAlaLeuGly 718
|||||
1691 GCGGTTATCTGATTAAACAGTGCATAGATTAAATTCGTAATATTAAATA 1740
```

```
719 TyrSerGluGlnProAspValGlu.....ProAspThrLeuG1 731
|||||
1741 GAGAAATCTATGATGATAAAAGAAATACGTCAATGGGAAGACGTTGCC 1790
731 nLysGlnAlaAspHisCysAsnAspArgArgMet.....Alas 744
|||||
1791 AGAATTAGCTGAGCACACATCACACGTGAACGCCGTGCCATTGAAGCGG 1840
744 erLysArgValGlnGluLeuSerIleGlyLeuPhePheAlaValLeuVal 760
|||||
1841 AACGTGTACTGATGAATTGAAAAAGCTGAGTAT.....ATGATT 1881
761 LysGluSerGlyProLeuGluSerGluAlaMetValMetGlyValLeuAs 777
|||||
1882 CAACATATTGTTGAT...GAATTTGAAGGTATCATTAGTCGTTGCTAA 1928
777 n 777
1929 t 1929
seq_name: /SIBSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH54223
seq_documentation_block:
ID AAH54223 standard; DNA; 3111 BP.
XX
AC AAH54223;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3587.
XX
KW Staphylococcus epidermidis SRI strain; Infection; diagnosis;
XX
KW vaccination; endocarditis; ds..
XX
OS Staphylococcus epidermidis.
XX
PN W0200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782..
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
Claim 8; Page 1177-1178; 2188pp; English.
XX
AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
```

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX  
 SQ Sequence 3111 BP; 1146 A; 485 C; 622 G; 858 T; 0 other;

# alignment\_scores:

Quality: 567.00 Length: 814  
 Ratio: 1.373 Gaps: 30  
 Percent Similarity: 50.737 Percent Identity: 25.061

## alignment\_block:

US-09-647-377-9 x AAH54223 ..

Align seg 1/1 to: AAH54223 from: 1 to: 3111

```

4 ProAspTyrLysLeuAsnLeuArgSerProGlyThrProArgGlyValSe 20
|||||
3 CCTGACTATGAA.....CCCATGTCAGTATC 28

20 rSerValValGlyProSerAlaValGlyAlaSerProGlyAspLysLys 37
|||||
29 TGACTTT.....CAAGATCGCTTAGTTAAACAGTCCGAC..... 65

37 exLysAsnLysSerMetArgGlyLysLysSerIlePheGlu..... 51
|||||
66 .....TCATTAGAGATTAAATAAATACTCGTTGAATTAGAA 104

52 .....ThrTyrMetSerLysGI 57
|||||
105 CAGTCGTGTTTAATTGAACGTACAAGACAGACAGATATCAAGCTAAACA 154

57 uAspValSerGluGlyLeuLysArgGlyThrLeuIleGlnGlyValLeuA 74
|||||
155 ATCCAACAACAACAAATCAAAA.....CTAATCAAGGACGCTTAA 195

74 rGileAsnProLysLysPheHisGluAlaPheIle...ProSerProAsp 89
|||||
196 CTCAAAATAAAAAAGCGTTT.....GCTTCTTAAGACCTGAAGATGAC 239

90 GlyAspArgAspIlePheIleAspGlyValValAlaArgAsnArgAlaLe 106
|||||
240 GAGATGGATGATATTTTATCCACCACTAAATC...AATAGACATT 286

106 uAsnGlyAspLeuValValLysLeuLeuProGluAspGlnTrpLysA 123
|||||
287 AGATGGAGACTGTCATCGTGGAA..... 311

123 laValLysProGluSerAsnAspLysGluIleGluAlaThrTyrGluAla 139
|||||
311 ..... 311

140 AspIleProGluGluGlyCysGlyHisProLeuGlnGlnSerArgLy 156
|||||
312 .....ATCAAAAATCTCGT... 326

156 sGlyTrpSerGlyProAspValIleIleGluAlaGlnPheAspSera 173
|||||
326 ..... 326

173 spSerGluAspArgHisGlyAsnThrSerGlyLeuValAspGlyValLys 189
|||||
327 .....GGAGACATAAAGTAAATTTGAAGTGAAGTAAATCTATTGAA 371

190 LysLeuSerIleSerThrProAspArgGlyLysGluAspSerSerThrPr 206
|||||
372 AAGCATTACAGT.....ACACA 388

206 oValMet.....LysAspGluAsnThrProIleProG 217
|||||
389 AGTTGTTGGACGTATAGCAAGCAAGCATTTTGGTTTCGTATTACCG 438

217 lnAspThrArgGlyLeuSerGluLysSerLeuGlnLysSerAlaLysVal 233

```

```

439 ATGACAAACGTATTATGCAAGAT..... 461
|||||
234 ValTyrIleLeuGluLysLysHisSerArgAlaAlaThrGlyIleLeuLy 250
|||||
462 ATCTTTATACCTAAAGGACAAAATTTAGGTGCTGTAGATGGT..... 503
|||||
250 sLeuLeuAlaAspLysAsnSerAspLeuPheLysLysTyrAlaLeuPheS 267
|||||
503 ..... 503

267 erProSerAspHisArgValProArgIleTyrValProLeuLysAspCys 283
|||||
504 .....CATAAAGTA..... 512

284 ProGlnAspPheMetThrArgProLysAspPheAlaAsnThrLeuPheIl 300
|||||
513 .....TT 514

300 eCysArgIleIleAspTrpLysGluAspCysAsnPheAlaLeuGlyGlnL 317
|||||
515 AGTACAATTTACCAAGTATGCCGATAGTACTGACAATCCAGAAGCCACG 564
|||||
317 euAlaLysSerLeuGlyGlnAlaGlyGluIleGluProGluThrGluGly 333
|||||
565 TCTCACAATATTAGTCTATAAATGATCCAGGTGTAGATATACTTCC 614
|||||
334 lLeuLeuThrGluTyrGlyValAspPheSerAspPheSerSerGluValLe 350
|||||
615 ATCATTTACCAAGTGAATAGAAATC...GAGTTTCCAGATGATGATT 661
|||||
350 uGluCysLeuProGlnSerLeuProTrpThrIleProProAspGluValG 367
|||||
662 ACAA...GAAGCTGAAGAGTACCGGATGTAATAGAACCTCTGAAATCG 708
|||||
367 lYlYsArgArgAspLeuArgLysAspCysIlePheThrIleAspProSer 383
|||||
709 AAGGCGTCTGTGATTTAAGAGATGAATTAACAATCACTATAGATGGCGCA 758
|||||
384 ThrAlaArgAspLeuAspAspAlaLeuAlaCysArgArgLeuThrAspGl 400
|||||
759 GATGCTAAAGATTAGATGATGCCATTGCTGTAATAAATAATTAATAAATGG 808
|||||
400 yThrPheGluValGlyValHisIleAlaAspValSerTyrPheValProG 417
|||||
809 CAACACCGAGCTTACAGTTAGTATTGTCAGATGTAAGTTACTATGTAAG 858
|||||
417 luGlySerSerLeuAspLysValAlaAlaGluArgAlaThrSerValTyr 433
|||||
859 AAGGATCAGCTTAGATAAAGAAGCTTATGATCGTCGACAAAGTGTGTAT 908
|||||
434 LeuValGlnLysValValProMetLeuProArgLeuLeuCysGluGluLe 450
|||||
909 CTGTGTCGATCGAGTAATCCGATGATCCACACCGCTCTAAGTAATGGAAT 958
|||||
450 uCysSerLeuAsnProMetThrAspLysLeuThrPheSerValIleTrpL 467
|||||
959 ATGCTCAITTAATCCAGAAGAGATCGTTTAACATTAAAGTTGTCGAATGG 1008
|||||
467 ySLeuThrProGluGlyLysIleLeuGluGluTrpPheGlyArgThrIle 483
|||||
1009 AATAAATGAACGAGCGCAAGTTGTAACACATGAAATCTTTGATAGTGTGA 1058
|||||
484 lLeaArgSerCysThrLysLeuSerTyrAspHisAlaGlnSerMetIleGl 500
|||||
1059 ATACATTCAACTACAGATGACATATGATGAGTAAACAAAATTATCAC 1108
|||||
500 uAsnProThrGluLysIleProGluGlu.....GluLeuProPro.... 513
|||||
1109 TGATCAAGATCTCTGAAATACGTTTCAACAATATAAGAGATTAAACACCTATGT 1158
|||||
514 .....lLeSerProGluHisSerValGluGluValHisGlnAlaValLeu 528
|||||

```

```
1159 TAGATTAGCGCAAGATTTTATCAATAGATTAAATTCAT..... 1196
529 AsnLeuHisSerIleAlaLysGlnLeuArgArgGlnArgPheValAspG1 545
1197 .....ATGCGCAACGTCGT.....GG 1213
545 yAlaLeuArgLeuAspGlnLeuLysLeuAlaPheThrLeuAspHisGluT 562
1214 AGAAATTGATTGTAT...ATTAATGAAGCGAAAGTACTTGTGAATGACG 1260
562 hrGlyLeuProGlnGlyCysHisIleTyrGluTyrArgAspSerAsnLys 578
1261 AAGGTATTCACACAGAAGTCTAATGAGAGAACGTCGCGGAAGGAGACGT 1310
579 LeuValGluGluPheMetLeuLeuAlaAsnMetAlaValAlaHisLysI1 595
1311 TTAATTGAATCATTCATGTTAGCAGCAATGAACAGTAGCTGAACAC... 1358
595 ePheArgThrPheProGluGlnAlaLeuLeuArgArgHisProProG 612
1359 .TTCATAAATTGGAAGTACCATTATTATTCGTGTTTCATGAACACCAA 1407
612 InThrLysMetLeuSerAspLeuValGluPheCysAspGlnMetGlyLeu 628
1408 AAPCTGACCGATTAAAGACAGTCTTCGACTTTATTACCAATTCGGTATT 1457
629 ProMetAspValSerSerAlaGlyAlaLeuAsnLysSerLeuThrLysTh 645
1458 .....ATCATAAAGG 1468
645 rPheGlyAspAspLysTyrSerLeuAlaArgLysGluValLeuThrAsnM 662
1469 TACAGTGAAGATATTTCACCAACATTCACAAAGCATTCACGAAGAAG 1518
662 eTyrSerArgProMetGlnMetAlaLeuTyrPheCysSerClyMetLeu 678
1519 TTGAAGGTGAGTACTACTACTACTACTACTACTACTACTACTACTACTA 1562
679 GlnAspGlnGluGln.....PheArgHisTyrAl 688
1563 CGTTTCATCAACAGCACATATGATGATGTTAATTAGGACATTTTGG 1612
688 aLeuAsnValProLeuTyrThrHisPheThrSerProIleArgArgPheA 705
1613 TTGTCTGTGAGTACTACTACTACTACTACTACTACTACTACTACTACTA 1662
705 laAspValIleValHisArgLeuLeuAlaAlaLeuGlyTyrSerGlu 721
1663 CTGATTAAACAGTCATAGATTAAATTCGTAATATTAAATAGAGAATTC 1712
722 GlnProAspValGlu.....ProAspThrLeuGlnLysGlnAl 734
1713 ATGGATAAAAGAAATACGTCATCTGGGAAGAGACGTTGCCAGAAATAGC 1762
734 aAspHisCysAsnAspArgArgMet.....AlaSerLysArgV 747
1763 TGAGCACACATCACACACGTGAACCGCGTCGTCATTGAAGCGCAACGTGA 1812
747 alGlnGluLeuSerIleGlyLeuPheAlaValLeuValLysGluSer 763
1813 CTGATGAATTGAAAAGCTGAGTAT.....ATGATTCAACATATT 1853
764 GlyProLeuGluSerGluAlaMetValMetGlyValLeuAsn 777
1854 GGATGAT...GAATTTGAAGGTATCATTAGCTCGTGTGCTAAT 1892
seq_name: /SIDSL/9c9data/hold-geneseq/geneseq-emb1/NA1999.DAT.AAX13264
seq_documentation_block:
ID AAX13264 standard; DNA; 6005 BP.
XX
AC AAX13264;
XX
```

```
DT 19-MAR-1999 (first entry)
XX Enterococcus faecalis genome contig SEQ ID NO:327.
DE
XX Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX Enterococcus faecalis.
OS
XX Enterococcus faecalis.
PN WO9850555-A2.
XX
XX 12-NOV-1998.
XX
XX 04-MAY-1998; 98WO-US08985.
XX
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Dillon PJ, Kunsch CA;
PI
XX WPI; 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 1424-1427; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 6005 BP; 2076 A; 973 C; 1287 G; 1659 T; 10 other;
SQ
alignment_scores:
Quality: 566.50 Length: 842
Ratio: 1.333 Gaps: 31
Percent Similarity: 50.475 Percent Identity: 24.466
alignment_block:
US-09-647-377-9 x AAX13264 ..
Align seg 1/1 to: AAX13264 from: 1 to: 6005
42 MetArgGlyLysLysLysSerIlePheGluThrTyrMetSerLysGluAs 58
|||:|||||:|||||: |||
955 ATGGAAGCCAAAGAAAGAAAGCTTT.....TCAATGGAAGA 992
|||:|||||:|||||: |||
58 pValSerGluGlyLeu..... 63
|||:|||||:|||||: |||
993 AATTCACAAGGCTTGAATTTAGAAAAGATTCAGATTTTAAATTTAG 1042
64 .....Lys 64
|||
1043 TGCACCAACCATTCACAAATGGAACGAGAAAGATCAGTTAGTTTCAACAA 1092
65 ArgGly.....ThrLeuIleGlnGlyValle 73
```

1093 uAGGCCAAAGTCTCTTACCAATGAAGACTTATTAAATGAAGAACGTT 1142  
 73 uArgIleAsnProLysLysPheHisGluAlaPheIleProSerProAspG 90  
 1143 TCGTGAATGAACGAGGCTTT.....GGTTTTGTAACCATCGATCCTG 1186  
 90 LyAspArgAspIlePheIleAspGlyValValAlaAlaArgAspArgAlaLeu 106  
 1187 AAGACACAGATGTTTATATCCG...AAGAGCCAAACGAACTTTCGAATG 1233  
 107 AsnGlyAspLeuValValLysLeuLeuProGlu...AspGlnTrpLy 122  
 1234 GATGGAGATACGGTTTAAATCGAGCTATCCACATCGGATCCCTTTTTC 1283  
 122 sAlaValLysProGluSerAsnAspLysGluIleGluAlaThrTyGluA 139  
 1284 AGATCGCGCCGACGAGGTAAGTCAAGAAATTAAGACGACGAGTGA 1333  
 139 laAspIleProGluGluGluCysGlyHisProluGlnGlnSerArg 155  
 1334 GCCAAGTTGTCGGAGAATTTGTGGCATATAGTGAAGAAGAAATGCGCAGAA 1383  
 156 LysGlyTrpSerGlyProAspValIleIleGluAlaGlnPheAspSe 172  
 1384 ATGGGACTGTGCTACATGATT.....CCCAA 1412  
 172 rAspSerGluAspArgHisGlyAsnThrSerGlyLeuValAspGlyValL 189  
 1413 AGATAAGAAATGGAATCAGTATCTACTGTATCAATTCACCTGAAGGGATTA 1462  
 189 ySLysLeuSerIleSerThrProAspArgGlyLysGluAspSerSerThr 205  
 1463 AGCCAGTA.....GATGGTAGCATT 1482  
 206 ProValMetLysAspGluAsnThrProIleProGlnAspThrArgGlyLe 222  
 1483 GTCATT..... 1488  
 222 uSerGluLysSerLeuGlnLysSerAlaLysValValTyrlleLeuGluL 239  
 1489 .....GCTGAAATTAACCTATTATCCAGATC 1513  
 239 ySLysHisSerArgAlaAlaThrGlyIleLeuLys...LeuLeuAlaAsp 254  
 1514 AAGAATATCCACAGATATGGAAGGACTAGTCAACAAGATGATTGGTCAT 1563  
 255 LysAsnSerAspLeuPheLysLysTyrrAlaLeuPheSerProSerAspHi 271  
 1564 AAAAAAT..... 1569  
 271 sArgValProArgIleTyrrValProLeuLysAspCysProGlnAspPheM 288  
 1570 .....GATCCAGAAATGGATATT 1588  
 288 etThrArgProLysAspPheAlaAsnThrLeuPheIleCysArgIleile 304  
 1589 TATCA..... 1593  
 305 AspTrpLysGluAspCysAsnPheAlaLeuGlyGlnLeuAlaLysSerLe 321  
 1593 ..... 1593  
 321 uClyGlnAlaClyGluIleGluProGluThrGluGlyIleLeuThrGluT 338  
 1594 .....ATCGTGGTGCTC 1606  
 338 yRGlyValAspPheSerAspPheSerSerGluValLeuGluCysLeuPro 354  
 1607 ATGGAATTCCC...ACAGCATTTTCGTATGAAGTTTGGCTGACGACGGAC 1653  
 355 GlnSerLeuProTrpThrIleProProAspGluValGlyLysArgArgAs 371  
 1654 CAA...GTACCAGAAACTATTGCAGAAACGATTTAGTCGGTCGTCGGGA 1700











```
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1912
US-08-487-890A-110

Query Match 1.3%; Score 38.8; DB 1; Length 1974;
Best Local Similarity 57.4%; Pred. No. 0.77;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 2774 tgctggcttttaggaatagacaccttttgacacacaaaggggatttttaatttggttttaa 2833
Db 1453 TGCAGGATGTCGGCTTGGCAGTACGGTGACCTAATAAGATTGATAATAGTCTTGAT 1394

QY 2834 caactcaaggggtttgttttttttttttttttttttttttttttttttttttttttttt 2893
Db 1393 AGATGTTGCGCTTGTTTCTTCTGCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 1334

QY 2894 tt 2895
Db 1333 TT 1332

RESULT 6
US-08-478-435-110/c
; Sequence 110, Application US/08478435
; Patent No. 5922323
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
```

```
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1912
US-08-478-435-110

Query Match 1.3%; Score 38.8; DB 2; Length 1974;
Best Local Similarity 57.4%; Pred. No. 0.77;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 2774 tgctggcttttaggaatagacaccttttgacacacaaaggggatttttaatttggttttaa 2833
Db 1453 TGCAGGATGTCGGCTTGGCAGTACGGTGACCTAATAAGATTGATAATAGTCTTGAT 1394

QY 2834 caactcaaggggtttgttttttttttttttttttttttttttttttttttttttttttt 2893
Db 1393 AGATGTTGCGCTTGTTTCTTCTGCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 1334

QY 2894 tt 2895
Db 1333 TT 1332

RESULT 7
US-08-337-483-110/c
; Sequence 110, Application US/08337483
; Patent No. 5922562
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
```

```
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1912
;
US-08-337-483-110

Query Match 1.3%; Score 38.8; DB 2; Length 1974;
Best Local Similarity 57.4%; Pred. No. 0.77; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 52;

QY 2774 tctgtgcttttaggaataggaacctttgacaccaaaggaggatttttaattggttttaa 2833
    ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1453 TGCAGGTATGTCGGCCTTGGCAGTACGCTGACCTAATAAGAATTGATAATAAGTCTTGAT 1394

QY 2834 caactcagggttggttgtttttatttttatttttcttcttatttttgcagctcagtt 2893
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1393 AGATGTTGTCGCTGTGTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1334

QY 2894 tt 2895
    ||
Db 1333 TT 1332

RESULT 8
US-08-478-373-110/c
; Sequence 110, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
```

```
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1912
;
US-08-478-373-110

Query Match 1.3%; Score 38.8; DB 2; Length 1974;
Best Local Similarity 57.4%; Pred. No. 0.77; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 52;

QY 2774 tctgtgcttttaggaataggaacctttgacaccaaaggaggatttttaattggttttaa 2833
    ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1453 TGCAGGTATGTCGGCCTTGGCAGTACGCTGACCTAATAAGAATTGATAATAAGTCTTGAT 1394

QY 2834 caactcagggttggttgtttttatttttatttttcttcttatttttgcagctcagtt 2893
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1393 AGATGTTGTCGCTGTGTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1334

QY 2894 tt 2895
    ||
Db 1333 TT 1332

RESULT 9
US-08-474-671-110/c
; Sequence 110, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
```



```
;
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 1038-720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1912
; US-08-897-438-110

Query Match 1.3%; Score 38.8; DB 4; Length 1974;
Best Local Similarity 57.4%; Pred. No. 0.77;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 2774 tgcgtgcttttaggaataggaacctttgacacaaaggagatttttaatttggttttaa 2833
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1453 TGCAGGTATGTCGGCCCTGGCAGTAGCGTGACCTAATAAGAAATTGATAAAGTCTTGAT 1394
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 2834 caactcagggttgggttttttttttttttttttttttttttttttttttttttttttt 2893
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1393 AGATGTTGCGCTTGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1334
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 2894 tt 2895
||
DB 1333 TT 1332

RESULT 12
US-08-637-654-110/c
; Sequence 110, Application US/08637654
; Patent No. 6358727
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Harkness, Robin E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew D
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:

;
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 1038-720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1912
; US-08-897-438-110

Query Match 1.3%; Score 38.8; DB 4; Length 1974;
Best Local Similarity 57.4%; Pred. No. 0.77;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 2774 tgcgtgcttttaggaataggaacctttgacacaaaggagatttttaatttggttttaa 2833
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1453 TGCAGGTATGTCGGCCCTGGCAGTAGCGTGACCTAATAAGAAATTGATAAAGTCTTGAT 1394
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 2834 caactcagggttgggttttttttttttttttttttttttttttttttttttttttttt 2893
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1393 AGATGTTGCGCTTGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1334
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 2894 tt 2895
||
DB 1333 TT 1332

RESULT 13
US-08-487-890A-108/c
; Sequence 108, Application US/08487890A
; Patent No. 5708149
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/487,890A
/ FILING DATE: 07-JUN-1993
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/175,116
/ FILING DATE: 29-DEC-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/148,968
/ FILING DATE: 08-NOV-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stewart, Michael I
/ REGISTRATION NUMBER: 24,973
/ REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (416) 595-1155
/ TELEFAX: (416) 595-1163
/ INFORMATION FOR SEQ ID NO: 108:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1993 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 3..1946
/ US-08-487-890A-108
```

```
Query Match 1.3%; Score 38.8; DB 1; Length 1993;
Best Local Similarity 56.2%; Pred. No. 0.77;
Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 2772 cctgctgcttttagaataagacccttttgacaccaaagggaatttttaattggtttt 2831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1480 CCTTAGGAATTTTCAGAACTGGGAGTAGCGAGACCTAATAAGATTGATAAGTGTC 1421

QY 2832 aacaactcaggggtgtgttttttttttttttttttttttttttttttttttttt 2891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1420 GATAGATTGTGCGCTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCTTCT 1361

QY 2892 tttttaaatg 2901
    ||| ||| ||
Db 1360 -TCYTAGGTG 1351
```

```
RESULT 14
US-08-478-435-108/c
; Sequence 108, Application US/08478435
; Patent No. 5922323
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/478,435
/ FILING DATE: 07-JUN-1995
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/337,483
/ FILING DATE: 08-NOV-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/175,116
/ FILING DATE: 29-DEC-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/148,968
/ FILING DATE: 08-NOV-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stewart, Michael I
/ REGISTRATION NUMBER: 24,973
/ REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (416) 595-1155
/ TELEFAX: (416) 595-1163
/ INFORMATION FOR SEQ ID NO: 108:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1993 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 3..1946
/ US-08-478-435-108
```

```
Query Match 1.3%; Score 38.8; DB 2; Length 1993;
Best Local Similarity 56.2%; Pred. No. 0.77;
Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 2772 cctgctgcttttagaataagacccttttgacaccaaagggaatttttaattggtttt 2831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1480 CCTTAGGAATTTTCAGAACTGGGAGTAGCGAGACCTAATAAGATTGATAAGTGTC 1421

QY 2832 aacaactcaggggtgtgttttttttttttttttttttttttttttttttttttt 2891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1420 GATAGATTGTGCGCTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCTTCT 1361

QY 2892 tttttaaatg 2901
    ||| ||| ||
Db 1360 TCYTAGGTG 1351
```

```
RESULT 15
US-08-337-483-108/c
; Sequence 108, Application US/08337483
; Patent No. 5922562
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-410 MTS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1993 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..1946  
US-08-337-483-108

Query Match 1.3%; Score 38.8; DB 2; Length 1993;  
Best Local Similarity 56.2%; Pred. No. 0.77;  
Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 2772 cctgtgctttaggaataggacctttgacaccaaaggaggatttttaatttggtttt 2831  
||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 1480 CCTTTAGGATTTTCAGACTGGAGTACGGAGACCTAATAAGATTGATAAAGTGTTTC 1421  
QY 2832 aacaactcaggggtttgttttatttttttttttttttttttttttttttttttt 2891  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 1420 CATAGATTTCGCTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTCT 1361  
QY 2892 tttttaaag 2901  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 1360 TCTTTAGGTG 1351

Search completed: July 8, 2002, 08:00:20  
Job time: 147 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 08:38:58 ; Search time 68.48 Seconds  
(without alignments)  
11119.507 Million cell updates/sec

Title: US-09-647-377-8  
Perfect score: 3100  
Sequence: 1 cggcgccgcgcgcctccgg.....tgaaaaaaaaaaaaaaaaaa 3100

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/pdata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/pdata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/pdata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/pdata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/pdata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/pdata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	0.8	2852	4	US-09-063-950-1
2	23	0.7	1279	3	US-09-248-335-25
3	23	0.7	4673	1	US-07-638-431-1
4	23	0.7	4673	5	PCT-US92-00018-1
5	22	0.7	284	4	US-09-040-984-80
6	22	0.7	284	4	US-09-123-912-80
7	22	0.7	751	1	US-08-592-936B-10
8	22	0.7	751	1	US-08-788-928A-16
9	22	0.7	751	2	US-09-111-573-10
10	22	0.7	953	1	US-08-197-793-1
11	22	0.7	953	2	US-08-636-176-1
12	22	0.7	953	5	PCT-US95-01618-1
13	22	0.7	1120	3	US-08-884-324-2
14	22	0.7	1120	4	US-08-832-180-6
15	22	0.7	1198	4	US-08-849-751-3
16	22	0.7	1198	4	US-09-478-816-3
17	22	0.7	1267	1	US-08-140-797-1
18	22	0.7	1267	1	US-08-486-670A-1
19	22	0.7	1357	6	5340934-7
20	22	0.7	1622	4	US-09-232-191-12
21	22	0.7	1622	4	US-09-232-200-12
22	22	0.7	1622	4	US-09-232-137-12
23	22	0.7	1622	4	US-09-232-201-12
24	22	0.7	1688	2	US-08-439-814-2
25	22	0.7	1716	2	US-08-954-333-9
26	22	0.7	1747	1	US-08-202-389-1
27	22	0.7	1777	2	US-08-937-466-5

28	22	0.7	1777	2	US-09-172-528-5	Sequence 5, Appli
29	22	0.7	1777	3	US-09-318-199-5	Sequence 5, Appli
30	22	0.7	1777	4	US-09-503-579-5	Sequence 5, Appli
31	22	0.7	1779	4	US-09-323-427-1	Sequence 1, Appli
32	22	0.7	1779	4	US-09-323-427-2	Sequence 2, Appli
33	22	0.7	1949	2	US-08-937-466-3	Sequence 3, Appli
34	22	0.7	1949	2	US-09-172-528-3	Sequence 3, Appli
35	22	0.7	1949	3	US-09-318-199-3	Sequence 3, Appli
36	22	0.7	1949	4	US-09-503-579-3	Sequence 3, Appli
37	22	0.7	2090	2	US-08-439-814-1	Sequence 1, Appli
38	22	0.7	2096	3	US-09-008-481A-10	Sequence 10, Appl
39	22	0.7	2096	3	US-09-195-666A-16	Sequence 16, Appl
40	22	0.7	2096	4	US-09-309-592-10	Sequence 10, Appl
41	22	0.7	2200	4	US-08-993-088A-1	Sequence 1, Appli
42	22	0.7	2200	4	US-08-993-424B-1	Sequence 1, Appli
43	22	0.7	2782	2	US-08-937-466-1	Sequence 1, Appli
44	22	0.7	2782	2	US-09-172-528-1	Sequence 1, Appli
45	22	0.7	2782	3	US-09-318-199-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-063-950-1  
; Sequence 1, Application US/09063950C  
; Patent No. 6235085  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: NOVEL LRSG PROTEIN AND NUCLEIC ACID MOLECULES AND USES  
; FILE REFERENCE: MEI-019  
; CURRENT APPLICATION NUMBER: US/09/063,950C  
; CURRENT FILING DATE: 1998-04-21  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2852  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (160)..(2178)  
US-09-063-950-1

Query Match 0.8%; Score 25; DB 4; Length 2852;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3076 aagtgtagaaaaaa 3100  
|||||  
Db 2801 aagtgtagaaaaaa 2825

RESULT 2  
US-09-248-335-25  
; Sequence 25, Application US/09248335  
; Patent No. 6095504  
; GENERAL INFORMATION:  
; APPLICANT: MCGONIGLE, BRIAN  
; APPLICANT: O'KEEF, DANIEL  
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES  
; FILE REFERENCE: CL-1128-A  
; CURRENT APPLICATION NUMBER: US/09/248,335  
; CURRENT FILING DATE: 1999-02-10  
; EARLIER APPLICATION NUMBER: 08/924,759  
; EARLIER FILING DATE: 1997-September-05  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: Microsoft Word Version 7.0A  
; SEQ ID NO 25  
; LENGTH: 1279  
; TYPE: DNA

; ORGANISM: maize  
US-09-248-335-25

Query Match 0.7%; Score 23; DB 3; Length 1279;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 23; Conservative 0; Mismatches 0; Indels

QY 3078 gtgtgaaaaaaaaaaaaaaaa 3100  
 |||||  
 Db 1200 gtgtgaaaaaaaaaaaaaaaa 1222

```

RESULT      3
US-07-638-431-1
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khusmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
;

```

Query Match 0.7%; Score 23; DB 1; Length 4673;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 23; Conservative 0; Mismatches 0; Indels

**Qy**    2851   tttatttttatttttccttttat   2873  
         | | | | | | | | | | | | | | | |  
**Db**    4437   TTTATTTTATTTCCTTTAT   4459

```

1  RESULT      4
2  PCT-US92-00018-1
3  Sequence 1, Application PC/TUS9200018
4  GENERAL INFORMATION:
5  APPLICANT: Hoffman, Stephen L.
6  APPLICANT: Charcoenvit, Yupin
7  APPLICANT: Hedstrom, Richard
8  APPLICANT: Khusmith, Srisin
9  APPLICANT: Rogers IV, William O.
10 TITLE OF INVENTION: Protective malaria sporozoite surface protein
11 TITLE OF INVENTION: immunogen and gene encoding
12 NUMBER OF SEQUENCES: 2
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: A. David Spevack
15 STREET: NMRDC Building 1 T-12 National Naval
16 STREET: Medical Center
17 CITY: Bethesda
18 STATE: MD
19 COUNTRY: USA
20 ZIP: 20814-5044
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: PatentIn Release #1.24
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: PCT/US92/00018
28 FILING DATE: 19921013
29 CLASSIFICATION: 424

```

Query Match 0.7%; score 23; DB 5; Length 4673;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 23; Conservative 0; Mismatches 0; Indels

Qy 2851 tttatttttatattttcccttttat 2873

```
Db 4437 TTTATTTTATTTTTCCTTTAT 4459
|||||
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (283)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-80

Query Match 0.7%; Score 22; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3079 tgtgaaaaaaaaaaaaaaaa 3100
|||||
Db 256 TGTGAAAAAAAAAAAAAAAA 277
|||||

RESULT 5
US-09-040-984-80
; Sequence 80, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEX:
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 284 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-040-984-80

Query Match 0.7%; Score 22; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3079 tgtgaaaaaaaaaaaaaaaa 3100
|||||
Db 256 TGTGAAAAAAAAAAAAAAAA 277
|||||

RESULT 6
US-09-123-912-80
; Sequence 80, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123,912A
; CURRENT FILING DATE: 1998-07-27
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 284
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (283)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-80

Query Match 0.7%; Score 22; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3079 tgtgaaaaaaaaaaaaaaaa 3100
|||||
Db 256 TGTGAAAAAAAAAAAAAAAA 277
|||||

RESULT 7
US-08-592-936B-10
; Sequence 10, Application US/08592936B
; Patent No. 5783393
; GENERAL INFORMATION:
; APPLICANT: Kellogg, Jill A.
; APPLICANT: Bestwick, Richard K.
; TITLE OF INVENTION: PLANT ISSUE / STAGE SPECIFIC PROMOTERS FOR
; TITLE OF INVENTION: REGULATED EXPRESSION OF TRANSGENES IN PLANTS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Behlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,936B
; FILING DATE: 29-JAN-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Evans, Susan T.
; REGISTRATION NUMBER: 38,443
; REFERENCE/DOCKET NUMBER: 4257-0012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 324-0880
; TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 751 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: pAG301 insert, drul cdna clone
US-08-592-936B-10

Query Match 0.7%; Score 22; DB 1; Length 751;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3079 tgtgaaaaaaaaaaaaaaaa 3100
|||||
Db 714 TGTGAAAAAAAAAAAAAAAA 735
|||||
```

```
/ /
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/111,573
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,936
FILING DATE: 29-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Evans, Susan T.
REGISTRATION NUMBER: 38,443
REFERENCE/DOCKET NUMBER: 4257-0012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 751 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: pag301 insert, drul cDNA clone
US-09-111-573-10

Query Match          0.7%; Score 22; DB 2; Length 751;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches    22; Conservative   0; Mismatches     0; Indels      0

QY  3079 tgtgaaaaaaagaaaaaaaaaaa 3100
       |||||||
Db    714 TGCGAAAAAAAAAAAAAAAAGAA 735

RESULT 10
US-08-197-793-1
Sequence 1, Application US/08197793
Patent No. 5510461
GENERAL INFORMATION:
APPLICANT: Meuer, S.
APPLICANT: Schraven, B.
APPLICANT: Schoenhaut, D.
APPLICANT: Ratnofsky, S.
TITLE OF INVENTION: pp32: A Newly Identified CD45-Associated
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/197,793
FILING DATE:
CLASSIFICATION: 530
```